Modeling Convex Subsets of Points – Part 1

Maurice Queyranne

Sauder School of Business at UBC, and
CORE, Université Catholique de Louvain

based on recent and current joint
with several collaborators…

LNMB Conference, Lunteren
January 12-13, 2016

ORBEL 30 CONFERENCE

JANUARY 28-29, 2016

A 50 DAYS@CORE EVENT

KEYNOTE LECTURES

Eva K. LEE (Georgia Technology Institute)
Yurii NESTEROV (Université catholique de Louvain)
Laurence A. WOLSEY (Université catholique de Louvain)

VENUE

Université catholique de Louvain
AGORA Auditorium (Place Agora, B-1348 Louvain-la-Neuve, Belgium)

REGISTRATION

To register, visit: http://www.orbel.be/orbel30

CORE 1966 - 2016 BRIDGING GAPS

UCL Université catholique de Louvain

CORE@50 CONFERENCE

BRIDGING GAPS ▸ MAY 23-27, 2016

CALL FOR PAPERS

KEYNOTE LECTURES

Marc FLEURBAEY (Princeton)
Michel GOEMANS (MIT)
George NEMHAUSER (Georgia Tech)
Victor CHERNOZHUKOV (MIT)

VENUE

Université catholique de Louvain
Louvain-la-Neuve, Belgium

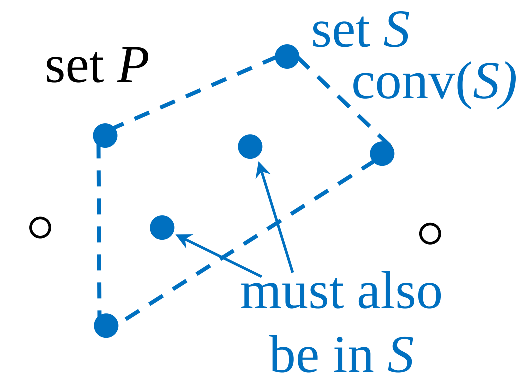
For information, to register or submit a paper:
http://www.core50.be/

UCL Université catholique de Louvain

CORE 1966 - 2016 BRIDGING GAPS

Convex Subsets of a Given Point Set

- A subset S of a given (finite) set P of n points in $\mathbf{R}^d$ is **convex** (relative to P) iff $S = P \cap \text{conv}(S)$ equivalently: $P \cap \text{conv}(S) \subseteq S$ (since $P \cap \text{conv}(S) \supseteq S$)
  - i.e., if we select a subset S of points in P then we must also select all points of P that are in their convex hull
  - definition also applies to more general *closure spaces* or *convexity spaces*



- We are interested in formulating the restriction "*the selected set of points must be convex*" in Integer Programming models
  - as hard or soft constraints

Why Convex Subsets?

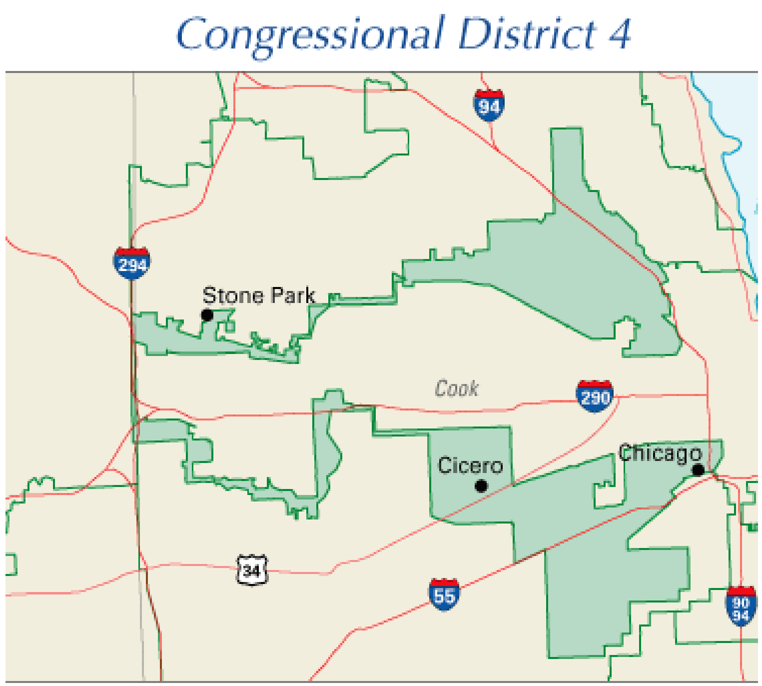
- Many (discrete) optimization models seek one or several subsets (of a given set of points, or of elementary regions, a.k.a., "cells") that should satisfy some "***shape constraint***"
  - often vaguely expressed: the set should "look compact", its shape should no be "too odd", etc.
- Convexity is one way of precisely formulating such shape constraints, which is appropriate (or approximately so) in some applications…

Why Convex Subsets? (2)

Some applications:

- Designing electoral districts
  - "Gerrymandering"

Gov. E. Gerry's "salamander"
Massachusetts, 1812



Illinois, 2004

- Douglas M. King, et al., 2012. Geo-Graphs: An Efficient Model for Enforcing Contiguity and Hole Constraints in Planar Graph Partitioning. *Operations Research* **60**(5) 1213–1228

Why Convex Subsets? (3)

More applications:

- Spatial Planning, e.g.,
  - Justin C. Williams, 2003. Convex land acquisition with zero - one programming. *Environment and Planning B: Planning and Design*, **30**, 255-270
- Farmland and Woodland Consolidation

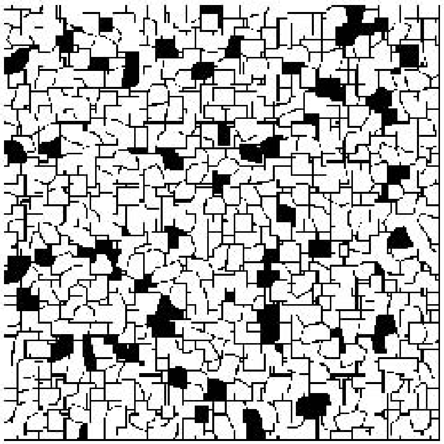
  (*Remembrement*, *Ruilverkaveling*)
  - Steffen Borgwardt, Andreas Brieden, and Peter Gritzmann, 2014. Geometric Clustering for the Consolidation of Farmland and Woodland. *Math. Intelligencer* **36** 37-44
- Police Quadrant Design
  - *work with Fernando Ordoñez and Flavio Guiñez (U. Chile)…*

Why Convex Subsets? (4)
Police Quadrant Design

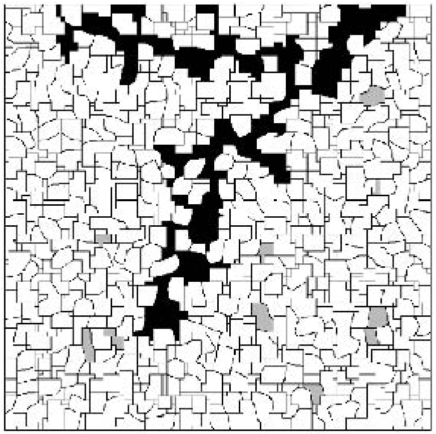
Why Convex Subsets? (5)

More applications:

- Forest Planning, e.g.,
  - M. Goycoolea, M., A. T. Murray, J.P. Vielma, A. Weintraub, 2009. Evaluating approaches for solving the area restriction model in harvest scheduling. *Forest Sci*. **55** (2) 149-165
  - Example: **old growth patch** from a Harvest Scheduling model
    - R. Carvajal, M. Constantino, M. Goycoolea, J.P. Vielma, A. Weintraub. Imposing Connectivity Constraints in Forest Planning Models, 2011



without shape constraints

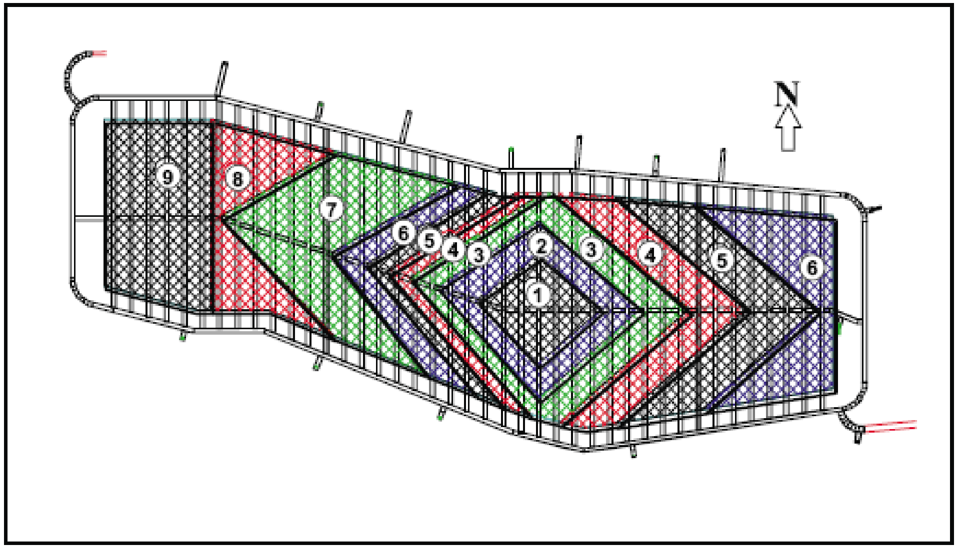


with connectivity constraint

Why Convex Subsets? (6)

More applications:

- Underground Mine Design and Scheduling, e.g.,
  - *PhD work of* ***Anita Parkinson*** *(UBC, 2012) with* ***Tony Diering*** *(Gemcom) and* ***Tom McCormick****, on drawpoint scheduling*
  - "convex" caves have better geomechanical stability
  - access constraints → 1D convexity along tunnels



current mine scheduling practice

Why Convex Subsets? (7)

Another 1-dimensional application:

- The Unit Commitment Problem in electric power generation
  - A generating unit must be "up" for successive time periods, i.e., for a convex subset of the planning horizon $P = \{1,2,\ldots,T\}$
  - Jon Lee, Janny Leung, François Margot, 2004. Min-up/min-down polytopes. *Discr. Opt.* **1** 77-85
  - Deepak Rajan, Samer Takriti, 2005. Minimum Up/Down Polytopes of the Unit Commitment Problem with Start-Up Costs. IBM Research Report.

Why Convex Subsets? (8)

…and more applications (in various dimensions):

- Data Mining, Statistical Clustering, Pattern Recognition, Data Compression, see, e.g., references in:
  - David Eppstein, Mark Overmars, Günter Rote, and Gerhard Woeginger, 1992. Finding Minimum Area k-gons. *Discrete and Computational Geometry* **7** 45-58
  - Paul Fischer, 1997. Sequential and parallel algorithms for finding a maximum convex polygon. *Computational Geometry* **7** 187-200.
  - C. Bautista-Santiago, J.M. Díaz-Báñez, D. Lara, P. Pérez-Lantero, J. Urrutia, and I. Ventura, 2011. Computing optimal islands. *Operations Research Letters* **39** (4) 246-251

Our Research Agenda

We seek to define notions of ("convex") shapes that are

- relevant to applications, and
- computationally tractable:
  - the Optimization Problem is efficiently solvable (or approximable)
  - the shape requirements can be enforced (or approximated) by a concise system of linear inequalities in natural and/or extended variables

Lectures Overview

Part 1: Computational complexity and algorithms

1. The Maximum Weight Convex Subset problem
2. Dimension 3 and higher: hardness results
3. One-dimension: a well understood case

Part 2: Modeling 2D and related convexities

1. 2D (points in the plane): DP algorithm for the optimization problem
2. 2D convex-shape constraints: IP modeling
3. Other notions of convexity
   a) Poset convexity
   b) Geodesic convexities and related notions

1. Maximum Weight Convex Subset Problem

To model convex-shape constraints with linear inequalities, the *Polytime Equivalence of Separation and Optimization* (GLS 1985) suggests considering the optimization problem:

- Given a set P of n points in $\mathbf{R}^d$ and **weights w_p** (of arbitrary sign) for all $p \in P$
- find a convex subset S of P with maximum weight $w(S) = \Sigma_{p \in S} w_p$

the **Maximum Weight Convex Subset problem**

- also of interest for Lagrangian Relaxation, and Column Generation

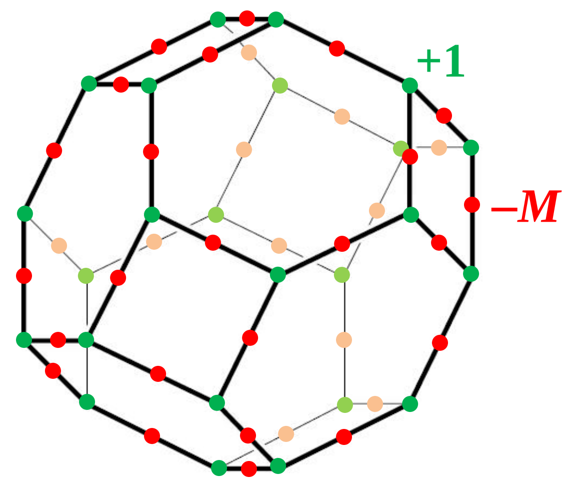
2. Dimension Three and Higher

- *Joint work with **Jeremy Barbay** (U.Chile), **Marek Chrobak** (U.C. Riverside) and **Miguel Constantino** (U. Lisboa)*

Theorem: The Maximum Weight Convex Subset problem is NP-hard for every dimension $d \geq 3$.

Proof: reduction from **MIS3CONPLAN**, the Maximum Independent Set problem on 3-connected planar graphs:

- embed instance $G = (V, E)$ of MIS3CONPLAN in $\mathbf{R}^3$ as the skeleton of a convex polytope with vertices $p(v)$ for all $v \in V$
- add the midpoint $p(u,v) = (p(u)+p(v))/2$ of every edge $(u,v) \in E$
- let weights $w_{p(u)} = 1$ and $$w_{p(u,v)} = -M < -|V|$$
- a subset S of these $|V|+|E|$ points is convex with $w(S) \geq 0$ iff $S \cap V$ is an independent set in G, and $w(S) = |S|$
- proof trivially extends to every dimension $d > 3$



QED

2. Dimension Three and Higher (2)

Theorem: When the dimension d is not fixed, the Maximum Weight Convex Subset problem cannot be approximated in polynomial time to within any factor $n^{1-\varepsilon}$ with $\varepsilon > 0$, unless P = NP

Proof: similar to previous proof, using the Maximum Independent Set (MIS) problem, known to be inapproximable to within any factor $n^{1-\varepsilon}$ with $\varepsilon > 0$, unless P = NP:

- embed instance $G = (V, E)$ of MIS in $\mathbf{R}^V$ where $p(v)$ is the v-th unit vector
- rest of the proof is identical
  - for every $(u,v) \in E$ add a point $p(u,v)$ at midpoint $(p(u)+p(v))/2$
  - let weights $w_{p(u)} = 1$ and $w_{p(u,v)} = -M < -|V|$
  - a subset S of these $|V| + |E|$ points is convex with $w(S) \geq 0$ iff $S \cap V$ is an independent set in G, and $w(S) = |S|$ **QED**

2. Dimension Three and Higher (3)

Theorem: The Maximum Weight Convex Subset problem is NP-hard for every dimension $d \geq 3$.

Theorem: When the dimension d is not fixed, the Maximum Weight Convex Subset problem cannot be approximated in polynomial time to within any factor $n^{1-\varepsilon}$ with $\varepsilon > 0$, unless P = NP

Open Problem: to find **approximation algorithms** (e.g., with constant factor, or PTAS, …) for the MWCS problem in dimension 3 (or higher)?

3. One-Dimension: a Well Understood Case

Given set P of n points $p_1 < p_2 < \ldots < p_n$ in $\mathbf{R}$, a subset $S \subseteq P$ is convex iff it consists of consecutive (or contiguous) points in P, i.e., iff $S = \{p_i, \ldots, p_j\}$ for some $i \leq j$ (or $S = \emptyset$)

- given any weights $w_1, \ldots, w_n$ the optimization problem (MWCS) is very easy
  - solved in linear time (by Dynamic Programming)
- with "natural" membership binary variables $y_j = 1$ iff $p_j \in S$ $O(n^3)$ **3-point constraints**

$$y_i \geq y_h + y_j - 1 \quad \text{for all } h < i < j$$

  suffice to enforce convexity of S (with binary variables)
  - Using $i = h+1$ suffices, so $O(n^2)$ constraints suffice

3. One-Dimension (2)

- … but the polyhedron defined by the 3-point alternating constraints and $0 \leq y \leq 1$ has fractional extreme points
- General **alternating constraints**:

$$\Sigma_{i=1,..,k} \ (-1)^{i+1} \ y_{d(i)} \ \leq \ 1 \ \text{ for every}$$

  - **odd** integer $k = 3, 5, \ldots \ (3 \leq k \leq n)$
  - (odd) subsequence $d(1) < d(2) < \ldots < d(k)$ of points

- Polyhedral result (Groeflin & Liebling, 1981; Lee & al., 2004):

  Theorem: In one dimension, the general alternating constraints and $0 \leq y \leq 1$ define the **convex hull** (in $\mathbf{R}^n$) of the set of all characteristic vectors y of contiguous subsets.

3. One-Dimension (3)

- General **alternating constraints:**

  $\Sigma_{i=1,..,k}\ (-1)^{i+1}\ y_{d(i)} \leq 1$ for every

  - **odd** integer $k = 3, 5, \ldots\ (3 \leq k \leq n)$ and
  - (odd) subsequence $d(1) < d(2) < \ldots < d(k)$
- About $2^{n\text{-}1}$ such constraints
- Cutting plane approach

  …requires solving the **Separation Problem**:

  Given vector y satisfying $0 \leq y \leq 1$

  - does y satisfy all alternating constraints?
  - and, if not, identify a violated constraint.
- O(n) exact separation algorithm (Lee & al.)

3. One-Dimension (4)

Extended Formulation:

- Auxiliary variable $F_i = 1$ indicates that point $p(i)$ is the **leftmost point** of the selected region S

$F_i \le y_i$	Point $p(i)$ must be selected to be leftmost
$F_i \ge y_i - y_{i-1}$	If $p(i)$ is selected but $p(i-1)$ is not, then $p(i)$ must be leftmost
$\sum_{i=1}^{n} F_i \le 1$	At most one leftmost point

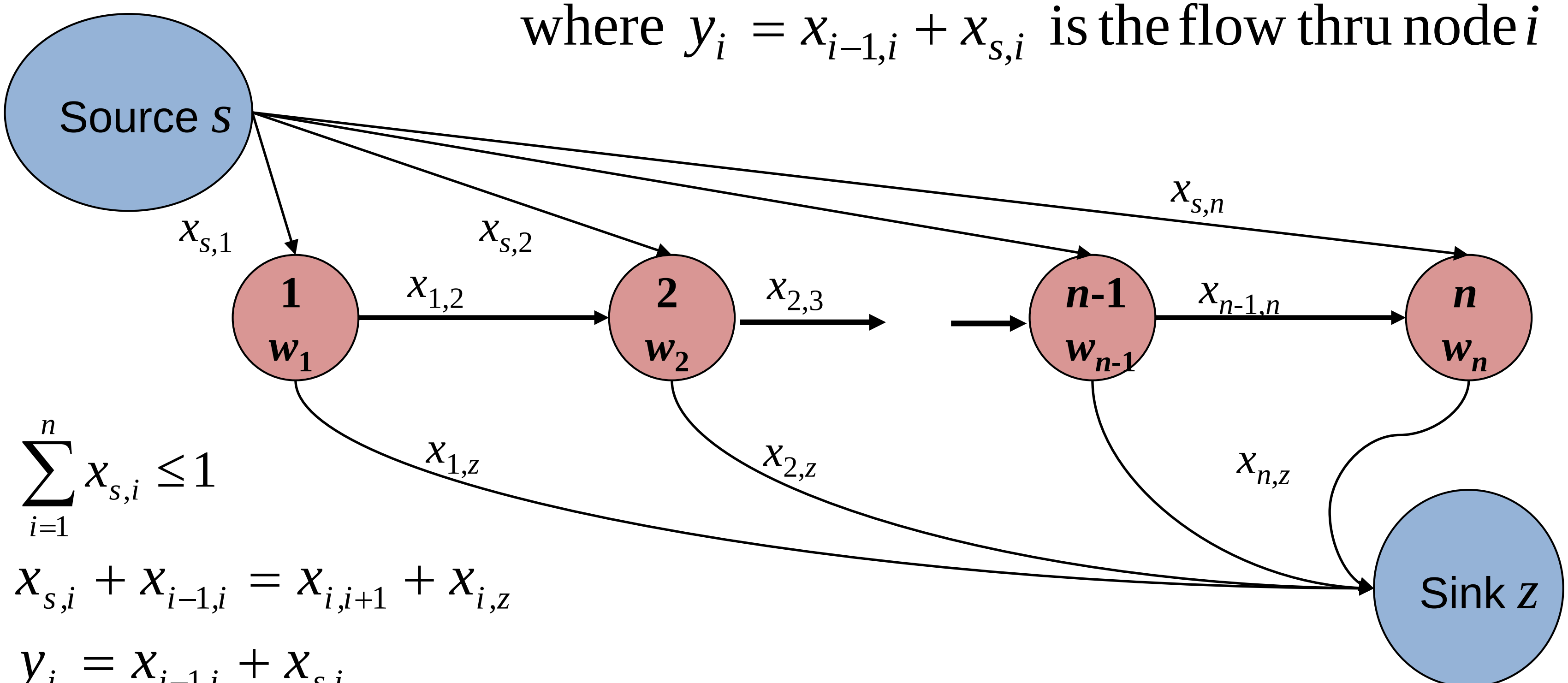
Theorem (Malkin & Wolsey, 2003; Rajan & Takriti,'05): This extended formulation is **ideal**, i.e., the extreme points of the corresponding polyhedron are the 0-1 vectors representing all the contiguous solutions

3. One-Dimension (5)

Formulation as a **Minimum Cost** (actually, Maximum Profit) **Network Flow Problem**:

- send at most 1 unit of flow to maximize $\sum_{i=1}^{n} w_i y_i$

where $y_i = x_{i-1,i} + x_{s,i}$ is the flow thru node i



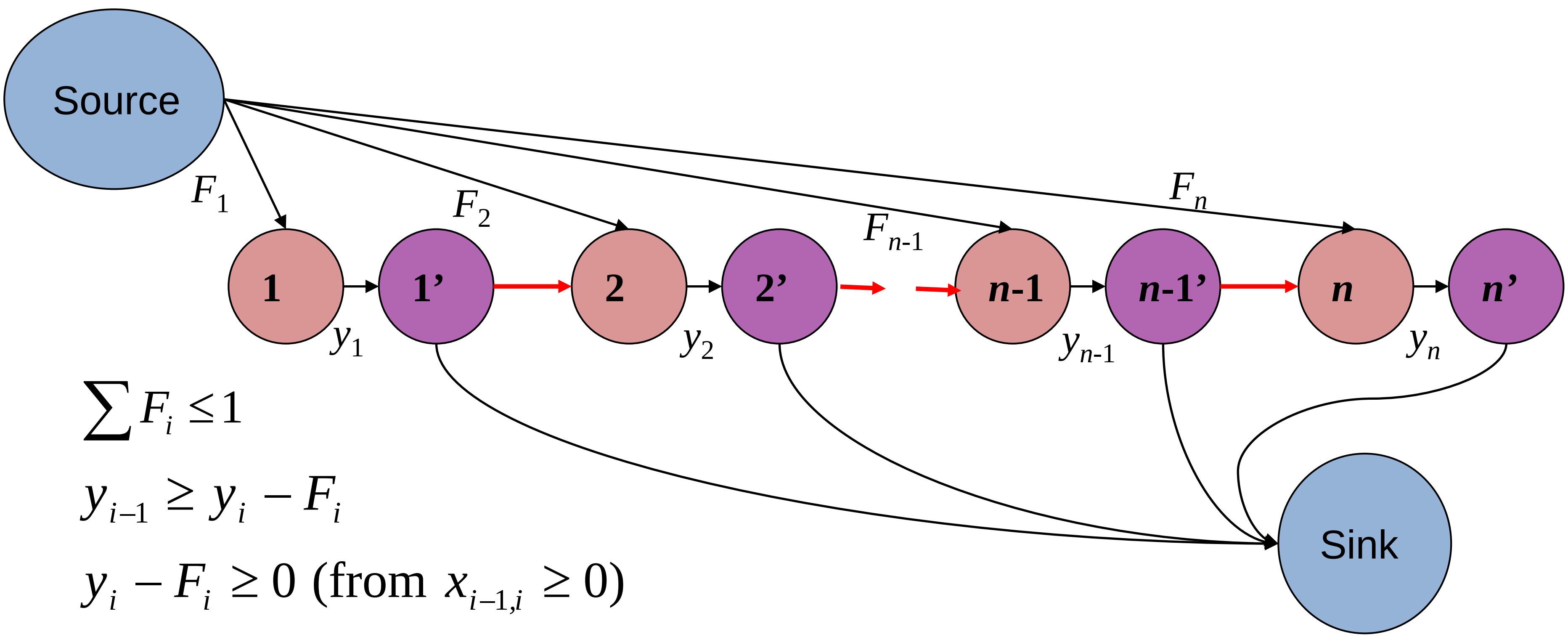
$$\sum_{i=1}^{n} x_{s,i} \leq 1$$

$$x_{s,i} + x_{i-1,i} = x_{i,i+1} + x_{i,z}$$

$$y_i = x_{i-1,i} + x_{s,i}$$

3. One-Dimension (6)

- Split each node to represent the flow through this node
- Extreme point solutions are integral
- Eliminate arc flow variables $x_{i\text{-}1,i}$ → Extended Formulation



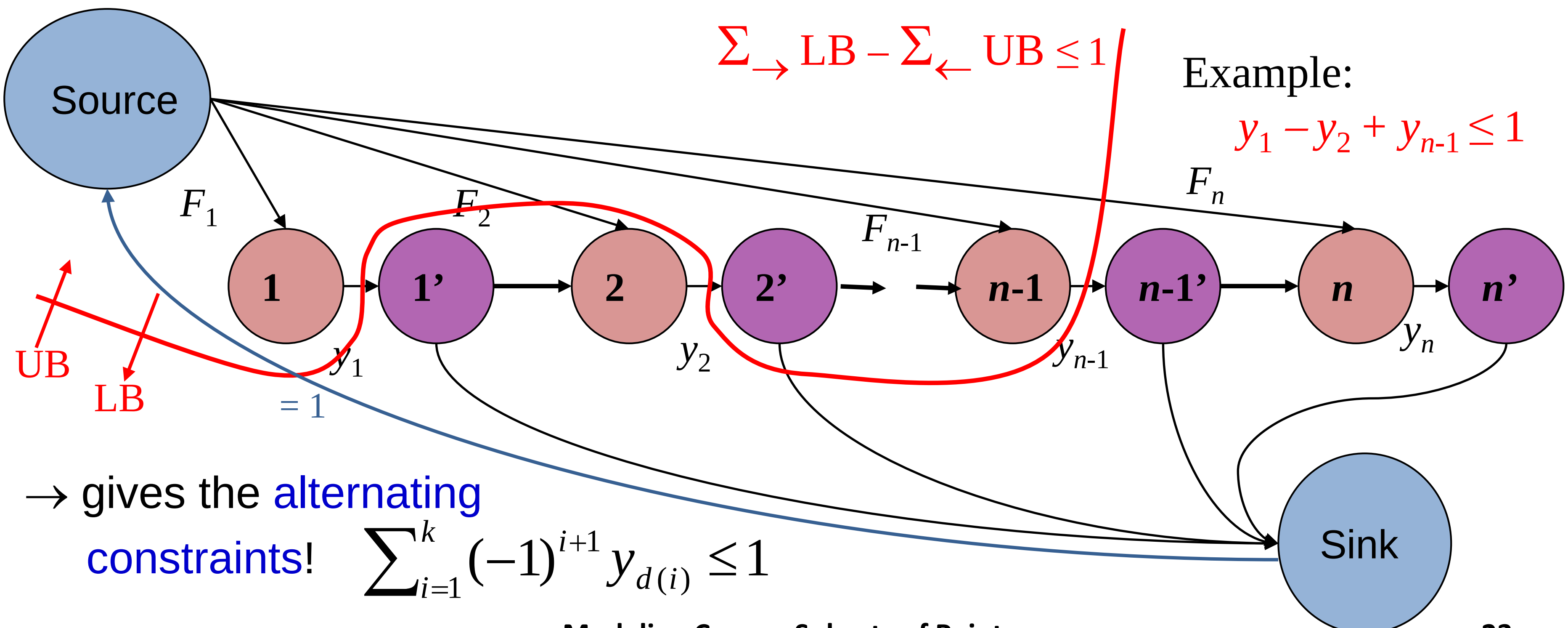
$$\sum F_i \leq 1$$

$$y_{i-1} \geq y_i - F_i$$

$$y_i - F_i \geq 0 \text{ (from } x_{i-1,i} \geq 0\text{)}$$

3. One-Dimension (7)

- Formulate Separation Problem for given vector y as feasibility flow with flow value 1; LB & UB y_i on split-node arcs, 0 & $+\infty$ elsewhere
- Apply Alan **Hoffman's *Feasible Circulation Theorem***:
  – finite capacity cuts $\leftrightarrow$ odd subsets $d(1) < d(2) < \ldots < d(k)$



→ gives the alternating constraints! $\sum_{i=1}^{k} (-1)^{i+1} y_{d(i)} \leq 1$

Lectures Overview

Part 1: Computational complexity and algorithms

1. The Maximum Weight Convex Subset problem
2. Dimension 3 and higher: hardness results
3. One-dimension: a well understood case

Part 2: Modeling 2D and related convexities

1. 2D (points in the plane): DP algorithm for the optimization problem
2. 2D convex-shape constraints: IP modeling
3. Other notions of convexity
   a) Poset convexity
   b) Geodesic convexities and related notions