Typical behavior of the linear programming method for combinatorial optimization problems: From a statistical-mechanical perspective

(J. Phys. Soc. Jpn. 83, 043801 (2014))

Satoshi Takabe

HUKUSHIMA Lab.
Graduate School of Arts and Sciences, The University of Tokyo

September 28, 2015

Introduction: Spin glass theory and its applications

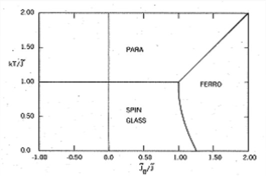
Spin glass theory

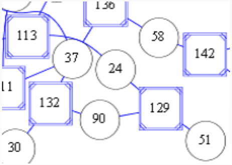
Study for diluted magnets.
Statistical mechanics for random systems,
a system with random interactions/fields

e.g.
$\mathcal{H} = \sum_{i,j} J_{ij}\sigma_i\sigma_j$,
$J_{ij} \sim P(J_{ij})$

Various applications

- Glasses
- Biological physics
- Theoretical computer science
  - Error correcting codes
  - Image restoration
  - Beysian inference
  - Constrainted satisfaction problems and
    Optimization problems





Optimization problems and statistical mechanics

Optimization problems

$$\textbf{Minimize} \quad f(\vec{x}) \ (\textbf{Cost function}),$$
$$\textbf{Subject to} \quad \vec{g}(\vec{x}) \geq 0,\ \vec{x} \in \chi^N \ (\textbf{Constraints})$$

Instance → (using an algorithm) → solution

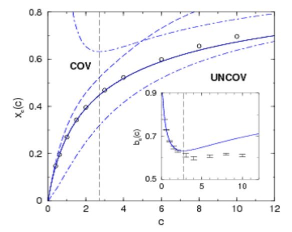
O.p. with discrete variables = Combinatorial Optimization Problem (COP)

Question

Can we estimate typical optimal values of randomized COPs?

Typical analyses by stat. mech.

- Randomized optimization problem
- Transform to randomized statistical-mechanical model.
- Estimate typical optimal value (averaged over random instances).



M. Weigt and A. K. Hartmann, Phys. Rev. Lett. **84**, 6118 (2000).

Today's goal

Approximation algorithms

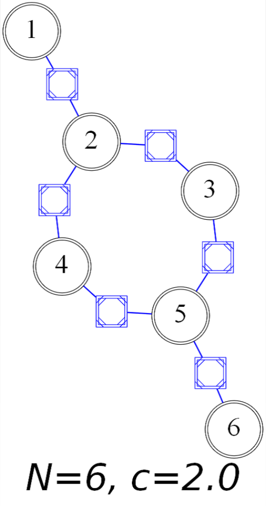
- COPs are generally NP-hard.
  $\longrightarrow$ Takes exp. time to solve COPs rigorously.
- Solve COPs in poly. time.
  $\longrightarrow$ Use approx. algorithms!
- In some cases, approx. algorithms perform well;
  they estimate optimal values with high accuracy.
- How well do they work typically? $\longrightarrow$ typical performance

Our goal

Analyze typical behavior of approximation algorithms for COPs
by using statistical-mechanics for random systems.

Graph theory

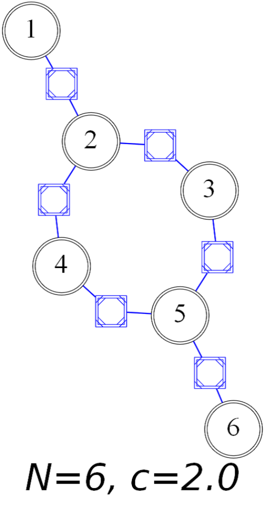
- Undirected (unweighted) graph $\boldsymbol{G = (V, E)}$
- $\boldsymbol{V}$: Vertex set, $|\boldsymbol{V}| = \boldsymbol{N}$
  e.g. $\boldsymbol{V = \{1, \cdots, 6\}}$
- $\boldsymbol{E \subset V^2}$: Edge set
  e.g. $\boldsymbol{E = \{(1,2), (2,3), \cdots\}}$
- Degree: # of edges connecting to a vertex
  e.g. Vertex 1 has degree 1
- Average degree $\boldsymbol{c}$: Average degree over vertices
- Cycle: a closed path
  e.g. there is a cycle with length 4



Minimum Vertex Cover problem (min-VC)

- Instance: Undirected graph $G = (V, E)$
- Cover or uncover each vertex.
- Cover all edges by covering vertices.
- An edge is covered if at least one connected vertex is covered.
- Minimize # of covered vertices.

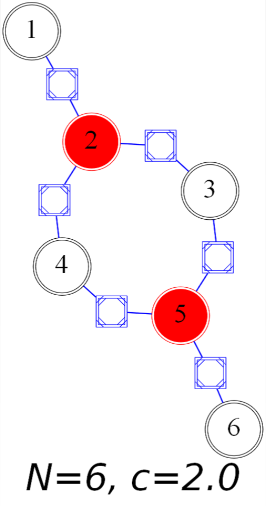
- A type of COPs
- Belongs to a class of NP-hard
- Application : Seeking a file on HDD, improving a group testing



Minimum Vertex Cover problem (min-VC)

- Instance: Undirected graph $G = (V, E)$
- Cover or uncover each vertex.
- Cover all edges by covering vertices.
- An edge is covered if at least one connected vertex is covered.
- Minimize # of covered vertices.

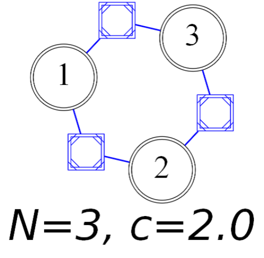
- A type of COPs
- Belongs to a class of NP-hard
- Application : Seeking a file on HDD, improving a group testing



Integer Programming problem (IP)

Formalization of min-VC:

- Assign a variable $x_i = \{0, 1\}$ to a vertex $i = \{1, \cdots, N\}$.
- $x_i = 1 \Leftrightarrow i$ is covered,
  $x_i = 0 \Leftrightarrow i$ is uncovered.
- Minimize # of covered vertices.
  $\longrightarrow$ Minimize $x_1 + x_2 + x_3$.
- Constraints: Cover all edges
  $x_1 + x_2 \geq 1,\ x_2 + x_3 \geq 1,\ x_3 + x_1 \geq 1,$
  $0 \leq x_i \leq 1,\ x_i \in Z\ (i = 1, 2, 3)$

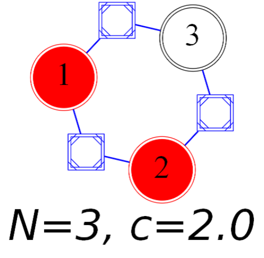


- COP with linear functions = Integer programming problem
- Optimal value: 2,
  Optimal solutions:$(x_1, x_2, x_3) = (1, 1, 0), (1, 0, 1), (0, 1, 1)$

Integer Programming problem (IP)

Formalization of min-VC:

- Assign a variable $x_i = \{0, 1\}$ to a vertex $i = \{1, \cdots, N\}$.
- $x_i = 1 \Leftrightarrow i$ is covered,
  $x_i = 0 \Leftrightarrow i$ is uncovered.
- Minimize # of covered vertices.
  $\longrightarrow$ Minimize $x_1 + x_2 + x_3$.
- Constraints: Cover all edges
  $x_1 + x_2 \geq 1,\ x_2 + x_3 \geq 1,\ x_3 + x_1 \geq 1,$
  $0 \leq x_i \leq 1,\ x_i \in Z\ (i = 1, 2, 3)$



- COP with linear functions = Integer programming problem
- Optimal value: 2,
  Optimal solutions:$(x_1, x_2, x_3) = (1, 1, 0), (1, 0, 1), (0, 1, 1)$

Relaxation to Linear Programming problem (LP)

IP: (generally) NP-hard
$\rightarrow$ Solve IP in poly. time (but approximately)

LP relaxation

Integer constraints of IP $x_i \in \mathbb{Z}$

Relax to real constraints $\Downarrow$ $x_i \in \mathbb{R}$

Linear Programming problem (LP)

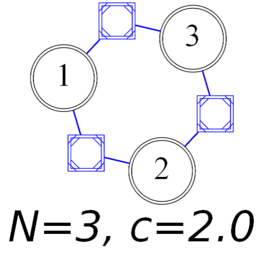
◯ Belongs to a class of P (solvable in poly. time).

△ LP optimal solutions differ from original IP.

Back to example

LP relaxation of min-VC:

$$\textbf{Minimize} \quad x_1 + x_2 + x_3$$
$$\textbf{Subject to}$$
$$x_1 + x_2 \geq 1,\ x_2 + x_3 \geq 1,\ x_3 + x_1 \geq 1,$$
$$0 \leq x_i \leq 1,\ x_i \in R\,(i = 1, 2, 3)$$



LP optimal value: 3/2,
LP optimal solution:
$$(x_1, x_2, x_3) = (1/2, 1/2, 1/2)$$

Questions

Is there a case where IP and its LP relaxed problem have the same optimal solutions?
What is the condition?
Otherwise, how do they differ?

Back to example

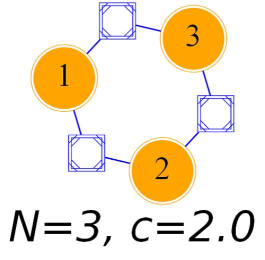
LP relaxation of min-VC:

Minimize $x_1 + x_2 + x_3$

Subject to

$x_1 + x_2 \geq 1,\ x_2 + x_3 \geq 1,\ x_3 + x_1 \geq 1,$

$0 \leq x_i \leq 1,\ x_i \in R\,(i = 1, 2, 3)$



LP optimal value: 3/2,
LP optimal solution:
$(x_1, x_2, x_3) = (1/2, 1/2, 1/2)$

Questions

Is there a case where IP and its LP relaxed problem have the same optimal solutions?
What is the condition?
Otherwise, how do they differ?

Hoffman-Kruskal's theorem

Mathematically rigorous result about IP and LP optimal solutions

Hoffman-Kruskal's theorem

A. J. Hoffman and J. B. Kruskal:
in "Linear Inequalities and Related Systems", pp. 223-246 (1956)

"Suppose an unweighted graph G,
G has no cycles with odd length.
$\Rightarrow$ IP and LP on G have same optimal solutions."

Other questions

What is a case where IP and LP have similarly the same optimal solutions in the order of N?
Otherwise, how do they differ?

IP and LP for min-VC

Instance: $G = (V, E)$

Normalize cost function.

Integer programming (IP)

Minimize

$N^{-1} \sum_i x_i$,

Subject to

$x_i + x_j \geq 1$ (if $(i, j) \in E$)

$0 \leq x_i \leq 1$, $x_i \in Z$.

Linear programming (LP)

Minimize

$N^{-1} \sum_i x_i$,

Subject to

$x_i + x_j \geq 1$ (if $(i, j) \in E$)

$0 \leq x_i \leq 1$, $x_i \in R$.

Algorithm: Simplex method (Danzig, 1947)

Difference between IP and LP

LP optimal solutions contain only $0, 1/2, 1$ (half-integrality; Nemhauser and Trotter, 1974).

IP and LP optimal solutions are coincident iff LP solution has no half-integer ($1/2$).

Typical analysis and random graphs

Definition of "similarly the same"

$\mathcal{G}$: graph ensemble with $N(\gg 1)$ vertices.

- x_c^{IP}: IP optimal value averaged over $\mathcal{G}$
- x_c^{LP}: LP optimal value averaged over $\mathcal{G}$

If

- $x_c^{\mathrm{IP}} = x_c^{\mathrm{LP}}$ and
- $o(N)$ half-integers in an LP optimal solution,

IP and LP have "similarly the same" optimal solutions

Erdös-Rényi random graph

One of basic graph ensembles.

1. Give a vertex set V.
2. Set edges with prob. p to each pair of vertices.
3. Parameter: average degree $c = 2p\,{}_N C_2/N \sim pN$

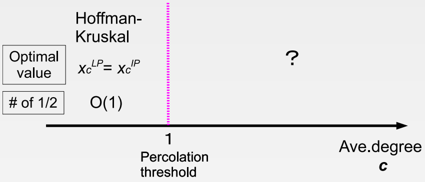
Suppose a graph is sparse; $c = O(N^0)$.

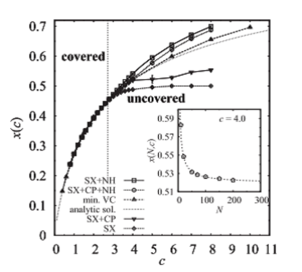
Typical analysis from H.K. theorem

From percolation theory of E.R. random graphs,
IP and LP have similarly the same optimal solutions if $c < 1$.
(Caution: sufficient condition)

$$\Downarrow$$

From numerical result they typically coincide up to
$c \sim 2.71 \simeq e$
T. Dewenter and A. K. Hartmann, Phys. Rev. E **86**, 041128 (2012)





Conjecture

IP and its LP relaxation for min-VC have similarily the same optimal solutions beyond percolation threshold $c = 1$.

Lattice gas model for min-VC

M. Weigt and A. K. Hartmann, Phys. Rev. Lett. **84**, 6118 (2000)

- Transformation: $\sigma_i = 2x_i - 1$
- 2-state model: $\sigma = \{\sigma_i\} = \{-1, 1\}^N$ ($\sigma_i = 1 \Leftrightarrow$ covered)
- Hamiltonian:

$$\mathcal{H}(\sigma) = \sum_i \sigma_i$$

- Grand canonical partition function:

$$\Xi = \sum_\sigma \exp(-\mu \mathcal{H}(\sigma)) \prod_{(i,j)\in E} \theta(\sigma_i + \sigma_j)$$

- IP optimal solutions = ground states
- Average optimal value x_c as $N \to \infty$,

$$x_c = \lim_{\mu\to\infty} \lim_{N\to\infty} \frac{1}{N} \mathsf{E} \left\langle \sum_i \sigma_i \right\rangle_\mu$$

  - **E**: Average over random graphs
  - $\langle \cdot \rangle_\mu$: Grand canonical ensemble average

Lattice gas model for LP with half-integrality

- Transformation: $\sigma_i = 2x_i - 1$, $x_i = \{0, 1/2, 1\}$
- 3-state model: $\sigma = \{\sigma_i\} = \{-1, 0, 1\}^N$ ($\sigma_i = 1 \Leftrightarrow$ covered)
- Hamiltonian

$$\mathcal{H}_r(\sigma) = \sum_i \sigma_i + \mu^{r-1} \sum_i (1 - \sigma_i^2)$$

  Second term: penalty term for $x_i = 1/2$
  parameter: $r \in \mathbf{R}$

- Grand canonical partition function:

$$\Xi = \sum_{\sigma} \exp(-\mu \mathcal{H}_r(\sigma)) \prod_{(i,j) \in E} \theta(\sigma_i + \sigma_j)$$

- Average optimal value x_c and average fraction of half-integers p_h as $N \to \infty$

Replica method

Random-averaged thermodynamic function $\mu J = -\lim_{N\to\infty} \mathsf{E}[\ln \Xi]$ is difficult to calcurate...

Replica method

Replica trick

$$\mathsf{E}[\ln \Xi] = \lim_{n\to 0} \frac{\mathsf{E}[\Xi^n] - 1}{n}$$

System σ is copied to n replicas σ^n.

- N spins to n spins: replicated vector $\vec{\xi}_i = (\sigma_i^1, \cdots, \sigma_i^n) \quad (i = 1, \cdots, N)$
- System is represented by $\{\vec{\xi}\}$.

Replica Symmetry (RS)

Order parameter and RS ansatz

- Order parameters: frequency dist. of $\vec{\xi}$: $c(\vec{\xi}) = N^{-1}\sum_i \delta_{\vec{\xi},\vec{\xi}_i}$
- Replica Symmetric (RS) ansatz:
  order parameter $c(\vec{\xi})$ is a function of $\xi \equiv \sum_{a=1}^{n} \xi^a$ and $\tilde{\xi} \equiv \sum_{a=1}^{n} (\xi^a)^2$.
- Laplace's transformation of $c(\vec{\xi})$

$$c(\vec{\xi}) \overset{\mathrm{RS}}{=} c(\xi, \tilde{\xi}) \equiv \int dP(h_1, h_2) Z^{-n} \exp(\mu h_1 \xi + \mu h_2 \tilde{\xi}),$$

$$Z = 1 + 2e^{\mu h_2}\cosh(\mu h_1)$$

  h_1: conjugate to ξ, h_2: conjugate to $\tilde{\xi}$
- Estimate $\mathbf{E}[\Xi^n]$ by saddle-point method under RS ansatz.

Saddle-point equations

Self-consistent equation of $P(h_1, h_2)$:

$$P(h_1,h_2) = \sum_{k=0}^{\infty} e^{-c}\frac{c^k}{k!}\int \prod_{i=1}^{k} dP(h_1^{(i)},h_2^{(i)})$$

$$\times\, \delta\left(h_1 + 1 + \sum_i u_2(h_1^{(i)},h_2^{(i)};\mu)\right)$$

$$\times\, \delta\left(h_2 - \mu^{r-1} + \sum_i [u_1(h_1^{(i)},h_2^{(i)};\mu) - u_2(h_1^{(i)},h_2^{(i)};\mu)]\right)$$

$$u_1(h_1,h_2;\mu) = \frac{1}{\mu}\ln[(1+\exp(\mu(h_1+h_2)))/Z],$$

$$u_2(h_1,h_2;\mu) = \frac{1}{2\mu}\ln[\exp(\mu(h_1+h_2))/Z],$$

$$Z = 1 + 2e^{\mu h_2}\cosh(\mu h_1)$$

3 large-μ limits with r

To analyze ground state, $\mu \to \infty$
Gibbs factor: $\exp[\mu \sum_i \sigma_i + \mu^r \sum_i (1 - \sigma_i^2)]$

- $r > 1$ IP-limit
  σ_i takes only ± 1 corresponding to IP optimal solution.
- $0 < r < 1$ LP-limit
  obtain LP optimal solution with minimum half-integers.
- $r \leq 0$ 3-state limit

IP-limit ($r > 1$)

$$P(h_1, \infty) = \sum_{k=0}^{\infty} e^{-c} \frac{c^k}{k!} \int \prod_{i=1}^{k} dP(h_1^{(i)}, \infty) \delta(h_1 + 1 + \sum_i \max(h_1, 0))$$

$h_2 \to \infty$ ($\mu \to \infty$) $\Leftrightarrow$ ground states without $\sigma_i = 0$ ($x_i = 1/2$)

RS solution of IP-limit

$$x_c^{\mathrm{IP}}(c) = 1 - \frac{W(c)^2 + 2W(c)}{2c}, \quad p_h(c) = 0$$

Lambert's W function: $W(x) \exp(W(x)) = x$

- RS: $c < e$, RSB (replica symmetry breaking): $c > e$ (e: Napier's number)

LP-limit ($0 < r < 1$)

$$x_c^{\mathrm{LP}}(c) = 1 - \frac{A + B + AB}{2c}, \quad p_h(c) = \frac{(B - A)(1 - A)}{c},$$

where A and $B(\geq A)$ obey $Ae^B = Be^A = c$.
RS solution is stable for any c.

The case of $c < e$

IP and LP have similarly the same solutions.

- Solution: $A = B$
- $x_c^{\mathrm{LP}}(c) = x_c^{\mathrm{IP}}(c)$
- $p_h(c) = 0$

The case of $c > e$

IP and LP have no common solutions.

- Solution: $A < B$
- $x_c^{\mathrm{LP}}(c) < x_c^{\mathrm{IP}}(c)$
- $p_h(c) > 0$

Numeral results

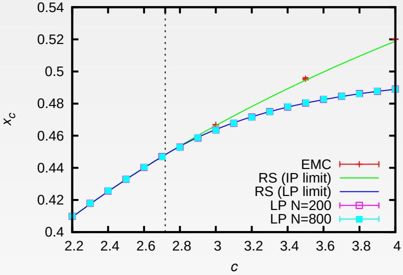


Figure : Average optimal value x_c as a function of average degree c.

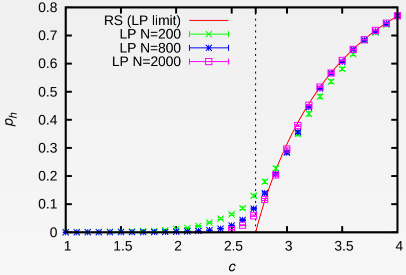


Figure : Average fraction of half-integers p_h as a function of average degree c.

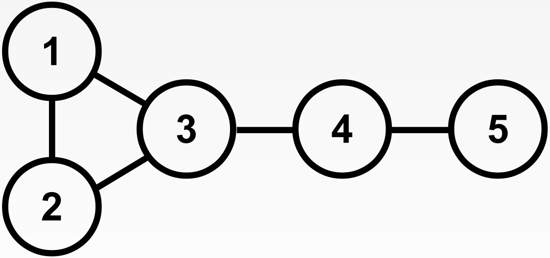
- Estimate x_c^{IP}: replica Exchange Monte Carlo method (EMC)
- Estimate x_c^{LP}: **lp_solve** (simplex algorithm)

Relation to other studies 1

Leaf Removal (LR) 1

A type of graph-removal algorithm:

- Repeat removing a leaf and connecting edges until there is no leaf.
  - Leaf: a pair of vertices $\{v, w\}$ where $(v, w) \in E$ and $deg(v) = 1$.

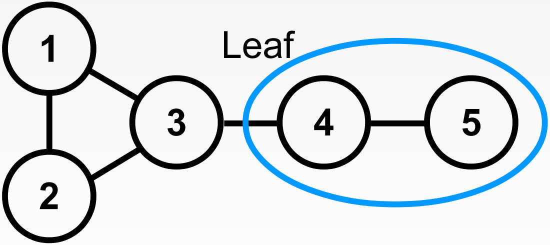


Relation to other studies 1

Leaf Removal (LR) 1

A type of graph-removal algorithm:

- Repeat removing a leaf and connecting edges until there is no leaf.
  - Leaf: a pair of vertices $\{v, w\}$ where $(v, w) \in E$ and $deg(v) = 1$.

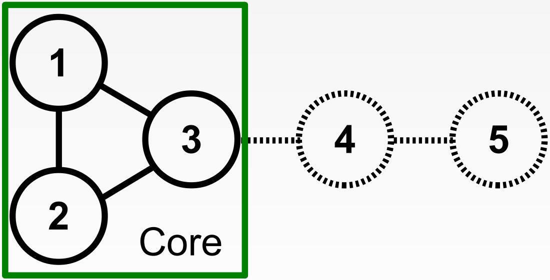


Relation to other studies 1

Leaf Removal (LR) 1

A type of graph-removal algorithm:

- Repeat removing a leaf and connecting edges until there is no leaf.
  - Leaf: a pair of vertices $\{v, w\}$ where $(v, w) \in E$ and $deg(v) = 1$.



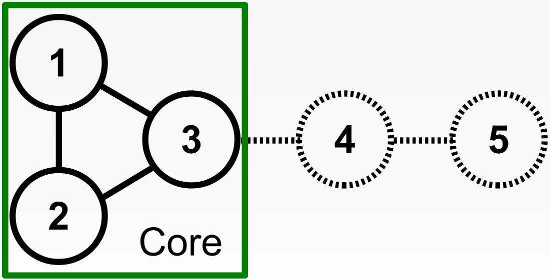
Relation to other studies 2

Leaf Removal (LR) 2

When LR stops,

- removed part: correctly assigned optimal variables
- core: connected components without leaves

If there is $O(N)$ core, LR cannot estimate optimal value x_c.
Otherwise, LR can estimate optimal value x_c correctly.



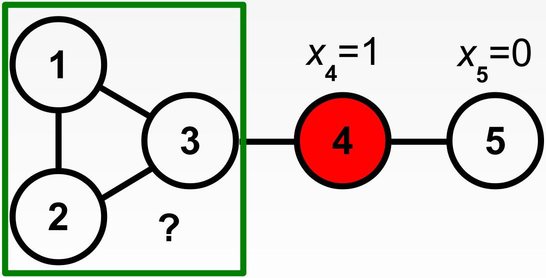
Relation to other studies 2

Leaf Removal (LR) 2

When LR stops,

- removed part: correctly assigned optimal variables
- core: connected components without leaves

If there is $O(N)$ core, LR cannot estimate optimal value x_c.
Otherwise, LR can estimate optimal value x_c correctly.



Relation to other studies 3

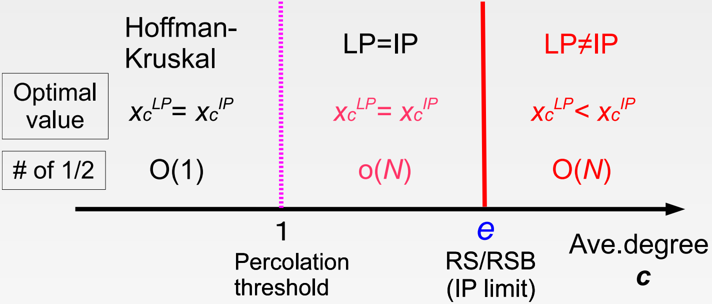
LP and LR core

Average fraction of half-integers p_h of LP
$=$ average LR core ratio $(N \to \infty)$
LR core: M. Bauer and O. Golinelli, Eur. Phys. J. B **24**, 339 (2001)

LR core makes min-VC difficult?

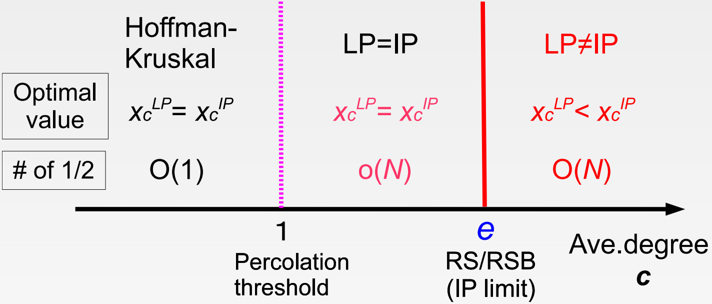
- $c < e$: no huge $(O(N))$ core $\to$ IP=LP=LR
  (good performance of approx. algorithms)
- $c > e$: huge core by LR
  - IP: splitting solution space (Barthel and Hartmann, 2004)
    $\to$ RSB (clustering)?
  - LP: assign variables to half-integers.
    $\to$ RS (one cluster)

Summary



- Statistical-mechanical analysis of typical behavior of LP for min-VC.
  - 3-state model reproducing IP and LP optimal solutions.
  - show a condition IP and LP have similarly the same optimal solutions as $N \to \infty$.
  - its threshold $c = e$ coincides with RS/RSB threshold of IP and is above percolation threshold $(c = 1)$.
  - Typical performance is related to other property (RS/RSB and LR core).

TRUE?



- Statistical-mechanical analysis of typical behavior of LP for min-VC.
  - 3-state model reproducing IP and LP optimal solutions.
  - show a condition IP and LP have similarly the same optimal solutions as $N \to \infty$.
  - its threshold $c = e$ coincides with RS/RSB threshold of IP and is above percolation threshold $(c = 1)$.
  - Typical performance is related to other property (RS/RSB and LR core) @03/2014

Thank you for your attention!

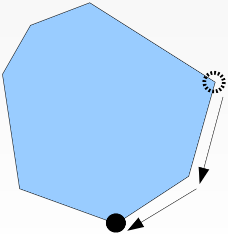
Algorithms for LP

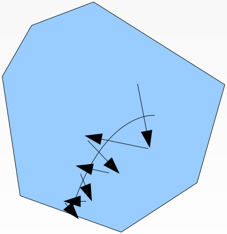
Simplex method

- Search extreme points of polytope.
- Worst case: takes exp. time
- Typical case: rapid

Interior method

- Search interior points of polytope
- Worst case: takes poly. time
- Typical case: sometimes more slowly than simplex method





Why do we need a penalty term?

Trivial ground states

Consider a simple example:

Minimize $x_1 + x_2$

Subject to $x_1 + x_2 \geq 1,\ 0 \leq x_i \leq 1,\ x_i \in R\,(i = 1, 2)$

Optimal solutions (or ground states)

- LP: $(x_1, x_2) = (1, 0), (0, 1)$ (see below)
- 3-state model without penalty term: $(x_1, x_2) = (1, 0), (1/2, 1/2), (0, 1)$

Simple 3-state model has trivial ground states (not LP optimal solutions).
$\rightarrow$ penalty term for $x_i = 1/2$