T-79.300 Postgraduate Course in Theoretical Computer Science:

Spectral Analysis of Landscapes

Petteri Kaski

Key ideas

▷ Landscape = configuration graph $G = (V, E)$
+ cost function $\psi : V \to \mathbb{R}$.

▷ The cost function ψ can be viewed as a vector in $\mathbb{R}^{|V|}$.

⇒ Does ψ have a "compact" representation in a more "suitable" basis of $\mathbb{R}^{|V|}$ (than the standard basis) ?

- Algebraic analysis of the configuration graph G helps here.
- The *Laplacian* of G gives a useful orthonormal eigenvector basis for $\mathbb{R}^{|V|}$.

▷ A landscape is *elementary* if its (zero-mean standardized) cost function $\bar{\psi}$ is an *eigenvector* of the Laplacian of G.

Key ideas

▷ Many relevant landscapes, such as $\{-1, 1\}$-spin glass [by definition], NAESAT [Grover, 1992], and TSP [Grover, 1992], are elementary.

▷ Elementary landscapes have a number of interesting properties

- all local cost maxima (resp. minima) lie at or above (resp. at or below) the average cost [Grover, 1992].
- a greedy local search starting from an arbitrarily "bad" configuration will reach an average cost configuration "quickly" [Grover, 1992].
- there is an upper bound on the number of *nodal domains* (maximal connected subgraphs of G where the sign of $\bar{\psi}$ is constant) [Davies *et al.*, 2001].

Outline

1. Preliminaries: Terminology for graphs, matrix theory related to graphs
2. Two relevant families of graphs and their spectra
   - ▷ the n-dimensional cube Q_2^n
     ("the" configuration graph for $\{-1, 1\}$-spin glass, NAESAT, ...)
   - ▷ the Cayley graph $\Gamma(S_n, T_n)$
     (a possible configuration graph for TSP and ASSIGNMENT)
3. Two landscapes: NAESAT is elementary. (TSP is elementary.)
4. Computing the Walsh decomposition of the NAESAT cost function (in polynomial time!) by sampling.
5. Grover's results on minima/maxima and behaviour of local search.
6. The discrete nodal domain theorem.

Terminology for graphs

- ▷ An *undirected graph* is a pair (V, E) where V is a nonempty set of *vertices* and E is a set of 2-subsets of V called *edges*.
- ▷ Two vertices $u, v \in V$ are *adjacent* if $\{u, v\} \in E$.
- ▷ A vertex $u \in V$ is *incident* to an edge $\{v, w\} \in E$ if $u \in \{v, w\}$.
- ▷ The number of vertices adjacent to a vertex is its *degree*.
- ▷ A graph is $(k$-$)$*regular* if all vertices have the same degree.

Adjacency and incidence matrices

▷ The *adjacency matrix* of $G = (V, E)$ is the $|V| \times |V|$ matrix A with rows and columns indexed by V:

$$A[u,v] = \begin{cases} 1 & \text{if } u, v \text{ are adjacent; or} \\ 0 & \text{otherwise.} \end{cases}$$

▷ The *incidence matrix* of $G = (V, E)$ is the $|V| \times |E|$ matrix D with rows indexed by V and columns indexed by E:

$$D[u,\{v,w\}] = \begin{cases} 1 & \text{if } u \in \{v, w\}\text{; or} \\ 0 & \text{otherwise.} \end{cases}$$

Matrix theory

- ▷ We work with finite dimensional vector spaces over $\mathbb{R}$.
- ▷ The *characteristic polynomial* of a real square matrix A is

$$p(\lambda) = \det(A - \lambda I).$$

- ▷ The zeros of $p(\lambda)$ are the *eigenvalues* of A.
- ▷ The *(algebraic) multiplicity* of an eigenvalue λ is the multiplicity of λ as a zero of p.
- ▷ An $n \times n$ real matrix has n eigenvalues, some of which may occur with multiplicity greater than one (and some of which may be complex).

Matrix theory (2)

▷ A vector $x \neq 0$ that satisfies the equation $(A - \lambda I)x = 0$ is called an *eigenvector* of A associated with $\lambda \in \mathbb{R}$.

▷ All eigenvectors associated with an eigenvalue λ together with the zero vector form a vector space, called the *eigenspace* of λ.

▷ The *(geometric) multiplicity* of an eigenvalue λ is the dimension of the corresponding eigenspace.

▷ The algebraic and geometric multiplicities of an eigenvalue may not agree. (It is even possible that a matrix has no eigenvectors.)

Matrix theory (3)

▷ The *inner product* of two column vectors $x, y \in \mathbb{R}^n$ is

$$\langle x, y \rangle \stackrel{\text{def}}{=} x^T y = \sum_{i=1}^{n} x_i y_i .$$

▷ The corresponding norm is $|x| = \langle x, x \rangle^{1/2}$.

▷ Two vectors x, y are *orthogonal* if $\langle x, y \rangle = 0$.

▷ A real square matrix A is *symmetric* if $A = A^T$.

Matrix theory (4)

Theorem. Let A be a real symmetric matrix. Then,

(i) the eigenvalues of A are real numbers; and

(ii) eigenvectors associated with distinct eigenvalues are orthogonal.

Theorem. Let A be a real symmetric $n \times n$ matrix. Then, $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of A.

Corollary. The algebraic and geometric multiplicities of eigenvalues agree for real symmetric matrices.

Matrix theory (5)

- ▷ The *spectrum* of a real square matrix A is a list of all of its eigenvalues together with their (algebraic) multiplicities.
- ▷ The *spectrum* of a graph G is the spectrum of its adjacency matrix A.
- ▷ The adjacency matrix is a real symmetric matrix.

Matrix theory (6)

A real symmetric $n \times n$ matrix A is *positive semidefinite* if $u^T Au \geq 0$ for all $u \in \mathbb{R}^n$.

Theorem. A real symmetric matrix A is positive semidefinite if and only if its eigenvalues are nonnegative.

Theorem. A real symmetric matrix A is positive semidefinite if and only if there exists a real matrix B such that $A = B^T B$.

Matrix theory (7)

- The *kernel* of a real $m \times n$ matrix A is the vector subspace of $\mathbb{R}^n$ that consists of all vectors $x \in \mathbb{R}^n$ such that $Ax = 0$.
- The *image* of a real $m \times n$ matrix A is the vector subspace of $\mathbb{R}^m$ that consists of all vectors $y \in \mathbb{R}^m$ such that $y = Ax$ for some $x \in \mathbb{R}^n$.
- The dimension of the image of A is denoted by $\mathrm{rk}\, A$.

Matrix theory (8)

Theorem. Let B be a real $m \times n$ matrix. Then,

$$\dim \operatorname{Ker} B + \operatorname{rk} B = n$$

Corollary. Let B be a real $m \times n$ matrix. Then,

(i) $\operatorname{Ker} B = \operatorname{Ker} B^T B$; and

(ii) $\operatorname{Ker} B^T = \operatorname{Ker} B B^T$; and

(iii) $\operatorname{rk} B = \operatorname{rk} B^T = \operatorname{rk} B^T B = \operatorname{rk} B B^T$.

The Laplacian of a (regular) graph

- ▷ Let $G = (V, E)$ be a k-regular graph, and suppose A is the adjacency matrix of G.
- ▷ Then, the *Laplacian* of G is the matrix

$$L = kI - A.$$

- ▷ The Laplacian L applied to a vector $x \in \mathbb{R}^{|V|}$ gives $y = Lx$, where

$$y_u = \sum_{\{u,v\} \in E} (x_u - x_v)$$

for all $u \in V$.

The spectrum of the Laplacian

Theorem. The Laplacian L of a graph G is a positive semidefinite matrix. Zero is always an eigenvalue of L with multiplicity c, where c is the number of connected components in G.

Theorem. If G is a k-regular graph, then the adjacency matrix A and the Laplacian $L = kI - A$ have identical eigenspace structure. The eigenspace of L associated with eigenvalue λ_L corresponds to the eigenspace of A associated with eigenvalue λ_A, where

$$\lambda_L = k - \lambda_A.$$

"**Corollary**." For k-regular graphs it suffices to study the spectrum of the adjacency matrix in order to study the spectrum of the Laplacian.

The n-cube Q_2^n.

- Denote by $\mathbb{Z}_2^n$ the set of all binary vectors $x = (x_1, x_2, \ldots, x_n)$ of length n.
- The *Hamming-distance* between two vectors $x, y \in \mathbb{Z}_2^n$ is

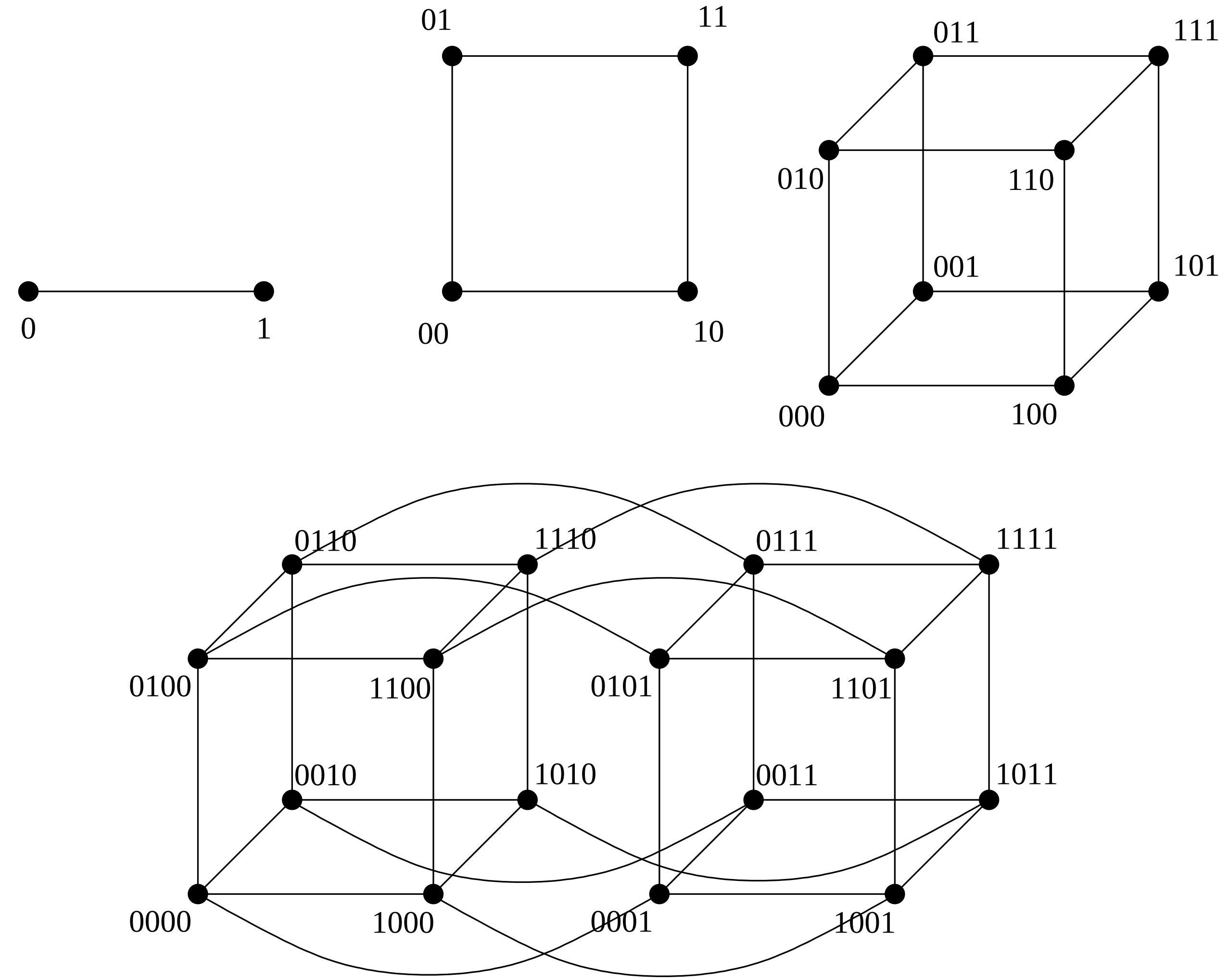
$$d_H(x, y) = |\{i \in \{1, \ldots, n\} \ : \ x_i \neq y_i\}|.$$

- The *Hamming weight* of a vector $x \in \mathbb{Z}_2^n$ is

$$w_H(x) = d_H(x, \vec{0}).$$

- The n-cube Q_2^n is the graph with vertex set $\mathbb{Z}_2^n$; two vertices $x, y \in \mathbb{Z}_2^n$ are connected by an edge if and only if $d_H(x, y) = 1$.

Example: Q_2^n for $n = 1, 2, 3, 4$



Properties of Q_2^n

Theorem. The n-cube Q_2^n is a connected bipartite n-regular graph with diameter n and order 2^n.

The spectrum of Q_2^n and the associated eigenspace structure is well-known. We proceed to describe it.

- ▷ Associate with each $z \in \mathbb{Z}_2^n$ a function

$$W_z : \mathbb{Z}_2^n \to \{-1, 1\},\ x \mapsto (-1)^{\sum_{i=1}^n x_i z_i}.$$

- ▷ The functions $\{W_z \ :\ z \in \mathbb{Z}_2^n\}$ are known as the *Walsh functions*.
- ▷ The *weight* of a Walsh function W_z is $w_H(z)$.

Example: The Walsh functions for $n = 3$

x	000	001	010	011	100	101	110	111
W_{000}	1	1	1	1	1	1	1	1
W_{001}	1	−1	1	−1	1	−1	1	−1
W_{010}	1	1	−1	−1	1	1	−1	−1
W_{011}	1	−1	−1	1	1	−1	−1	1
W_{100}	1	1	1	1	−1	−1	−1	−1
W_{101}	1	−1	1	−1	−1	1	−1	1
W_{110}	1	1	−1	−1	−1	−1	1	1
W_{111}	1	−1	−1	1	−1	1	1	−1

The spectrum and eigenspace structure of Q_2^n

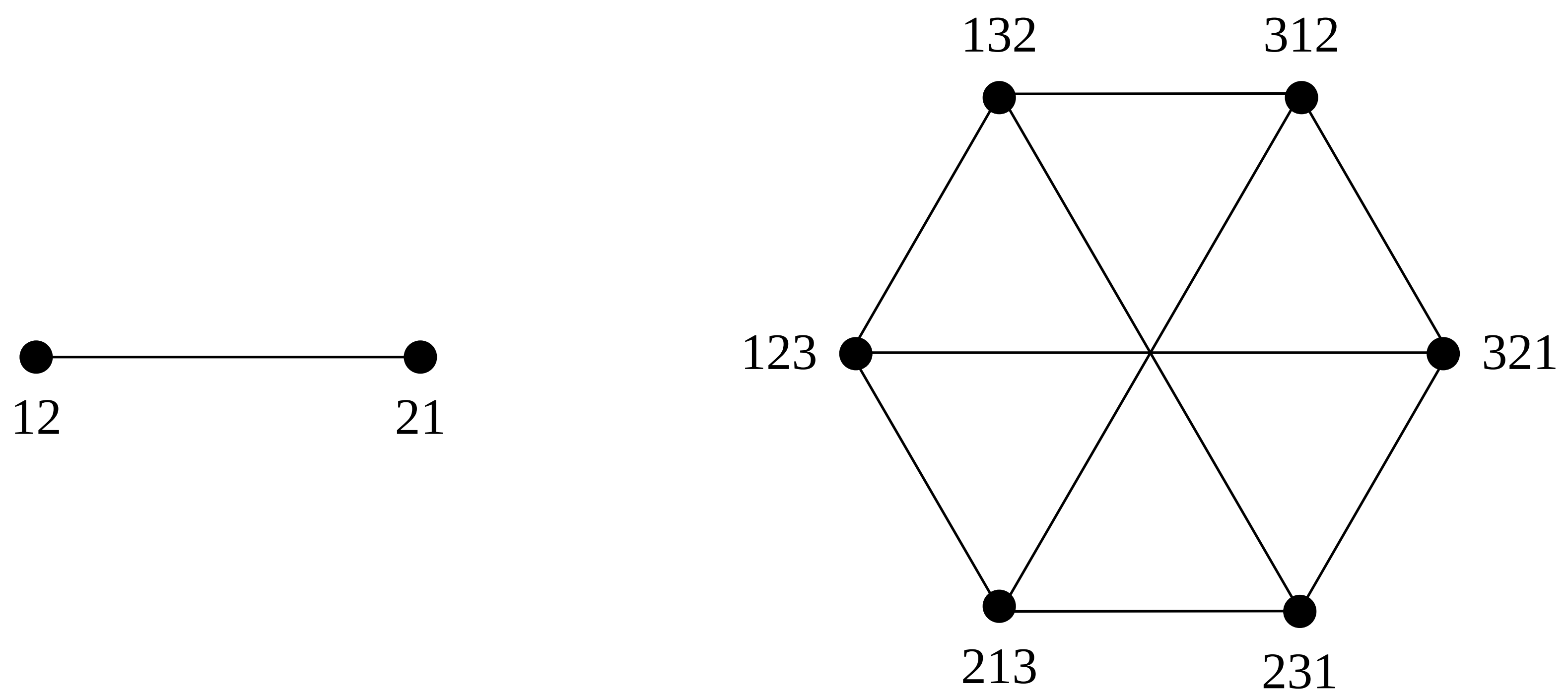
Theorem. The normalized Walsh functions $\tilde{W}_z = 2^{-n/2} W_z$ give an orthonormal basis of $\mathbb{R}^{2^n}$.

Theorem. Let A be the adjacency matrix of Q_2^n. Then, the distinct eigenvalues of A are $\lambda_k = n - 2k$, where $k = 0, 1, \ldots, n$, and λ_k has multiplicity $\binom{n}{k}$. An orthonormal basis for the eigenspace of λ_k is given by the weight k normalized Walsh functions.

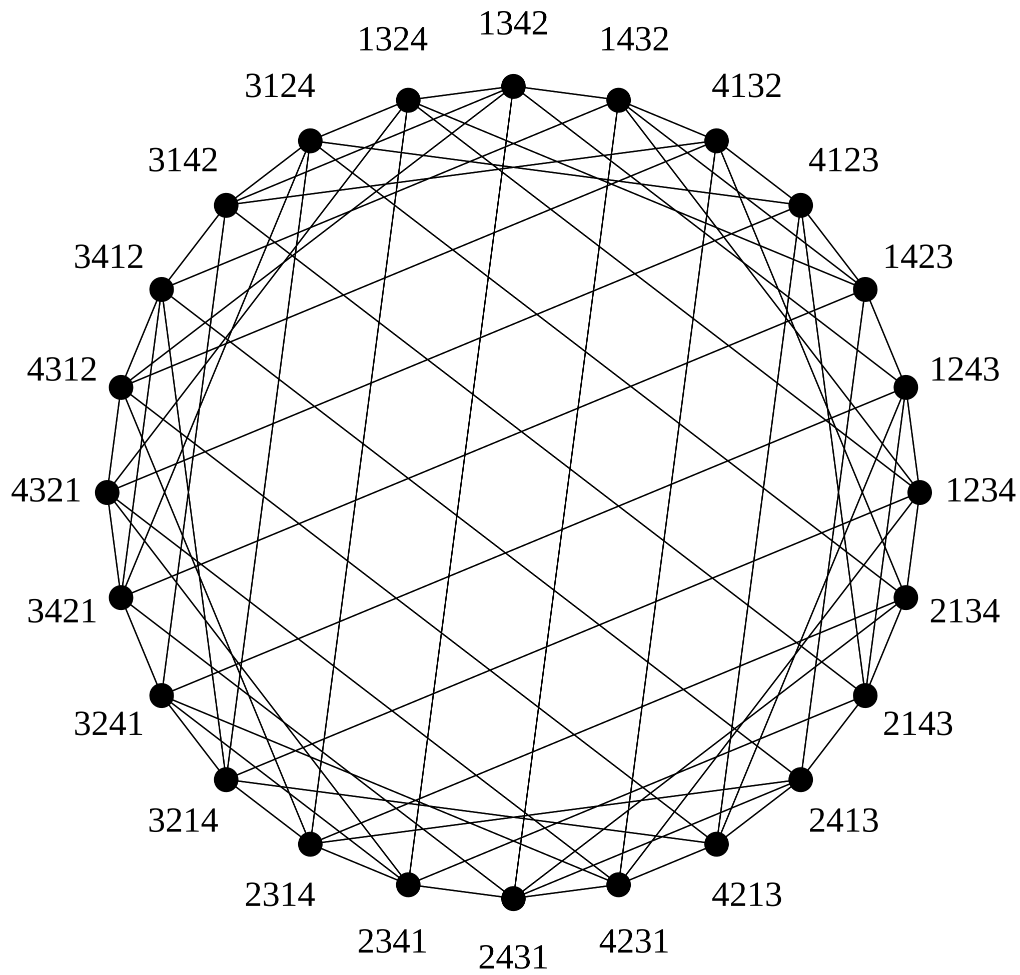
The Cayley graph $\Gamma(S_n, T_n)$

- ▷ A *permutation* of a nonempty set E is a bijection of E onto E.
- ▷ A *transposition* is a permutation that swaps two points and keeps the remaining points fixed.
- ▷ The *symmetric group* S_n on $\{0, 1, \ldots, n-1\}$ is the group formed by the set of permuations of $\{0, 1, \ldots, n-1\}$ with composition as the group operation.
- ▷ We denote the set of transpositions of the elements in $\{0, 1, \ldots, n-1\}$ by T_n.
- ▷ The Cayley graph $\Gamma(S_n, T_n)$ is the graph with vertex set S_n; two vertices $\pi_1, \pi_2 \in S_n$ are connected by an edge if and only if $\pi_2^{-1}\pi_1 \in T_n$.

Example: $\Gamma(S_n, T_n)$ for $n = 2, 3$



Example: $\Gamma(S_4, T_4)$



Properties of $\Gamma(S_n, T_n)$

- **Theorem**. The Cayley graph $\Gamma(S_n, T_n)$ is a connected bipartite $\binom{n}{2}$-regular graph with diameter $n - 1$ and order $n!$.
- The spectrum of $\Gamma(S_n, T_n)$ is known, however, its description requires an excursion into *representation theory* of the symmetric group, upon which we shall **not** embark here.
- The spectrum of $\Gamma(S_n, T_n)$ and the analysis of the TSP landscape is possibly (?) the topic of another presentation.

Not-all-equal satisfiability (NAESAT)

- Let $V = \{v_1, \ldots, v_n\}$ be a set of n binary variables.
- A *literal* is either a variable $v \in V$ or its complement $\bar{v}$.
- A *clause* is a set of three literals.
- A *truth assignment* is a vector $x \in \mathbb{Z}_2^n$, which assigns the value x_i to the variable v_i for all $i = 1, \ldots, n$.
- A truth assignment *satisfies* a clause if and only if all of its literals do not have the same truth value. The truth value of $\bar{v}$ is the opposite of the value assigned to v.

NAESAT

▷ The decision problem NOT-ALL-EQUAL SATISFIABILITY (NAESAT) asks, given a set C of clauses over n binary variables, whether there exists a truth assignment $x \in \mathbb{Z}_2^n$ to the variables that satisfies every clause in C.

▷ NAESAT is NP-complete.

▷ Associated landscape:

- Configuration graph: the n-cube Q_2^n
- Cost function: $\psi(x) = \#$ clauses **not** satisfied by $x \in \mathbb{Z}_2^n$

NAESAT is elementary

▷ Recall that a landscape is *elementary* if the (zero-mean standardized) cost function $\bar{\psi}$ is an eigenfunction of the configuration graph Laplacian.

▷ Proof strategy [Grover, 1992]:

- Compute the mean cost over all truth assignments $x \in \mathbb{Z}_2^n$.
- Establish that the zero-mean cost function $\bar{\psi}$ is an eigenfunction of the Laplacian of Q_2^n.

NAESAT is elementary (2)

Theorem.

$$\frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} \psi(x) = \frac{|C|}{4}.$$

Proof. Denote by $\psi_{C_i}(x)$ the cost of a single clause C_i from C. In other words,

$$\psi_{C_i}(x) = \begin{cases} 0 & \text{if } x \text{ satisfies } C_i\text{; and} \\ 1 & \text{otherwise.} \end{cases}$$

Since exactly 2 of the 8 possible truth value assignments to the three literals in a clause satisfy it, we have

$$\sum_{x \in \mathbb{Z}_2^n} \psi(x) = \sum_{i=1}^{|C|} \sum_{x \in \mathbb{Z}_2^n} \psi_{C_i}(x) = \sum_{i=1}^{|C|} 2^{n-3} \cdot 2 = |C| 2^{n-2}.$$

NAESAT is elementary (3)

Theorem. Define $\bar{\psi}(x) = \psi(x) - |C|/4$, and let L be the Laplacian of the n-cube Q_2^n. Then, $L\bar{\psi}(x) = 4\bar{\psi}(x)$.

Proof. By definition of the Laplacian on the n-cube

$$L\bar{\psi}(x) = \sum_{d_H(x,y)=1} \big(\bar{\psi}(x) - \bar{\psi}(y)\big) = \sum_{i=1}^{|C|} \sum_{d_H(x,y)=1} \big(\psi_{C_i}(x) - \psi_{C_i}(y)\big).$$

Now,

$$\sum_{d_H(x,y)=1} \big(\psi_{C_i}(x) - \psi_{C_i}(y)\big) = \left\{ \begin{array}{ll} -1 & \text{if } x \text{ satisfies } C_i\text{; and} \\ 3 & \text{otherwise,} \end{array} \right\} = 4\psi_{C_i}(x) - 1.$$

Thus,

$$L\bar{\psi}(x) = 4\psi(x) - |C| = 4\bar{\psi}(x).$$

NAESAT is elementary (4)

Corollary. The zero-mean cost function $\bar{\psi}$ is a vector in the eigenspace of L spanned by the $\binom{n}{2}$ weight 2 Walsh functions.

Corollary. For any NAESAT instance over n variables, the cost function ψ can be written as a sum of Walsh functions

$$\psi = \frac{|C|}{4} W_{\vec{0}} + \sum_{w_H(y)=2} \alpha_y W_y,$$

where $\alpha_y \in \mathbb{R}$. Thus, the number of clauses $|C|$ and the $\binom{n}{2}$ coefficients α_y completely characterize the cost function.

Computing α_y

- ▷ The standard way is to compute the *Walsh transform* (i.e. project ψ to each of the basis functions W_y separately). This, however, requires at least 2^n steps.
- ▷ A faster (polynomial in n) method exists that works by sampling ψ in $\binom{n}{2}$ locations followed by multiplication of the sample vector by a $\binom{n}{2} \times \binom{n}{2}$ matrix.
- ▷ Unfortunately, knowledge of the coefficients α_y does not (significantly) help in locating a minimum cost truth assignment:
- ▷ Locating an $x \in \mathbb{Z}_2^n$ that minimizes

$$\psi(x) = \alpha_{\vec{0}} W_{\vec{0}}(x) + \sum_{w_H(y)=2} \alpha_y W_y(x)$$

is *exactly* the $\{-1, 1\}$-spin glass optimization problem.

Grover's [Grover, 1992] result on local minima and maxima

▷ Fix an elementary landscape (G, ψ), and consider the normalized cost function $\bar{\psi}$.

▷ A configuration $u \in V$ is a *local minimum* if $\psi(u) \leq \psi(v)$ for all $\{u, v\} \in E$.

▷ In other words, $\psi(u) - \psi(v) = \bar{\psi}(u) - \bar{\psi}(v) \leq 0$ for all $\{u, v\} \in E$.

▷ Then by definition of the Laplacian L, $L\bar{\psi}(u) \leq 0$.

▷ By elementarity and positive definiteness of L, $L\bar{\psi}(u) = \lambda\bar{\psi}(u)$ for some $\lambda \geq 0$.
($\lambda = 0$ iff $\bar{\psi}$ is constant on each connected component of G.)

▷ Thus, $\bar{\psi}(u) \leq 0$ unless $\bar{\psi}$ is constant on every connected component of G.

Grover's result on greedy local search

- Let $L\bar{\psi} = \lambda\bar{\psi}$, $\lambda > 0$.
- Recall that $L\bar{\psi}(u) = \sum_{\{u,v\}\in E}\big(\bar{\psi}(u) - \bar{\psi}(v)\big)$.
- Thus, for every $u \in V$ such that $\bar{\psi}(u) \geq 0$, there exists a $\{u, v\} \in E$ such that

$$\bar{\psi}(u) - \bar{\psi}(v) \geq \frac{1}{\deg(u)} L\bar{\psi}(u) = \frac{1}{\deg(u)} \lambda\bar{\psi}(u).$$

- In other words,

$$\bar{\psi}(v) \leq \left(1 - \frac{\lambda}{\deg(u)}\right)\bar{\psi}(u) \leq \left(1 - \frac{\lambda}{\Delta}\right)\bar{\psi}(u)$$

Grover's result on greedy local search (2)

▷ Assuming that ψ takes only integral values, greedy local search will reach a configuration $w \in V$ with $\bar{\psi}(w) \leq 0$ in at most

$$m < \frac{1 - \log N}{\log\left(1 - \frac{\lambda}{\Delta}\right)}$$

steps if the zero-mean cost of the initial configuration is $N \geq 1$.

▷ Proof idea:

$$\left(1 - \frac{\lambda}{\Delta}\right)^m N < 1.$$

The discrete nodal domain theorem

- Let A be a real $n \times n$ symmetric matrix with nonpositive off-diagonal elements.
- Associate with A a graph Γ with vertex set $V = \{1, \ldots, n\}$ and edge set defined by $\{u, v\} \in E$ if and only if $A[u, v] < 0$.
- (In particular, we may take A to be the Laplacian of a graph G, in which case $G = \Gamma$.)

The discrete nodal domain theorem (2)

- ▷ Let $x \in \mathbb{R}^n$.
- ▷ A *strong positive (negative)* sign graph S is a maximal, connected subgraph of Γ whose vertices i satisfy $x_i > 0$ $(x_i < 0)$.
- ▷ A *weak positive (negative)* sign graph S is a maximal, connected subgraph of Γ whose vertices i satisfy $x_i \geq 0$ $(x_i \leq 0)$.

The discrete nodal domain theorem (3)

- ▷ Sort the eigenvalues of A to ascending order (first is smallest).
- ▷ Consider the mth eigenvalue λ_m of A that satisfies

$$\lambda_{m-1} < \lambda_m = \lambda_{m+1} \cdots = \lambda_{m+r-1} < \lambda_{m+r}.$$

Thus, the multiplicity of λ_m is r.

- ▷ Let $x \neq 0$ be an associated eigenvector, that is, $Ax = \lambda_m x$.
- ▷ Then, x has at most $m + r - 1$ **strong** sign graphs, and at most m **weak** sign graphs.

The discrete nodal domain theorem (4)

▷ For **any** NAESAT instance over n variables, the zero-mean cost function $\bar{\psi}$ has at most $1 + n + \binom{n}{2}$ **strong** sign graphs, and at most $1 + n + 1$ **weak** sign graphs.

▷ Proof idea:

- Put $A =$ the Laplacian of Q_2^n.
- $\bar{\psi}$ is in the eigenspace of eigenvalue 4, which has multiplicity $\binom{n}{2}$.
- eigenvalue 0 has multiplicity 1, eigenvalue 2 has multiplicity n.

▷ Unfortunately, this result does not imply anything about the number of local minima or maxima.