Penn State University
CSE 564 Complexity of Combinatorial Problems
Instructor: Pei Wu

Module
12

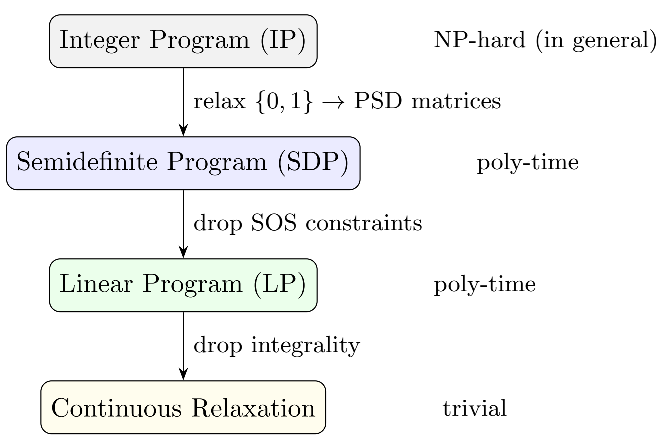
Max Cut and Goemans Williamson Algorithm

12.1 From Linear to Semidefinite Programming

12.1.1 A brief hierarchy of relaxations

Many combinatorial optimization problems can be written as integer programs: maximize a linear or polynomial objective subject to the constraint that variables take values in $\{0, 1\}$ or $\{+1, -1\}$. Since this is NP-hard in general, a standard strategy is to *relax* the integrality constraints and solve the relaxed problem in polynomial time, then *round* the fractional solution back to an integer solution.

The quality of the final integer solution depends critically on the *tightness* of the relaxation. We have a natural hierarchy:



SDPs occupy a sweet spot: they are strictly more powerful than LPs (they can capture non-linear geometric structure), yet still solvable in polynomial time.

12.1.2 Notation and prerequisites

We write $\mathcal{S}^n$ for the space of $n \times n$ real symmetric matrices and $\mathcal{S}^n_+$ for the cone of positive semidefinite (PSD) matrices. Recall:

$$X \succeq 0 \iff v^\top X v \geq 0 \; \forall v \in \mathbb{R}^n \iff \text{all eigenvalues of } X \text{ are } \geq 0.$$

The *trace inner product* on $\mathcal{S}^n$ is $\langle A,\, B\rangle := \mathrm{tr}(A^\top B) = \sum_{i,j} A_{ij} B_{ij}$.

12.2 Semidefinite Programs

12.2.1 Standard form

Definition 12.1 (Semidefinite Program). A *semidefinite program (SDP)* in standard (primal) form is:

$$\max_{X \in \mathcal{S}^n} \quad \langle C,\, X\rangle \quad \text{subject to} \quad \langle A_i,\, X\rangle = b_i \; (i = 1, \ldots, m), \quad X \succeq 0. \tag{12.1}$$

The data are: a symmetric cost matrix $C \in \mathcal{S}^n$, constraint matrices $A_1, \ldots, A_m \in \mathcal{S}^n$, and right-hand sides $b_1, \ldots, b_m \in \mathbb{R}$.

The single constraint $X \succeq 0$ is what distinguishes an SDP from an LP. If we replaced $X \succeq 0$ by "$X_{ij} \geq 0$ for all i, j", we would obtain a (larger) LP. Instead, $X \succeq 0$ is a *convex* constraint since $\mathcal{S}^n_+$ is a convex cone.

Remark 12.2. An LP is an SDP with X constrained to be *diagonal*: $X = \mathrm{diag}(x_1, \ldots, x_n)$, so $X \succeq 0$ reduces to $x_i \geq 0$.

12.2.2 The dual SDP

Definition 12.3 (SDP dual). The *dual* of (12.1) is

$$\min_{y \in \mathbb{R}^m} \quad b^\top y \quad \text{subject to} \quad \sum_{i=1}^{m} y_i A_i - C \succeq 0. \tag{12.2}$$

Theorem 12.4 (Weak and strong duality). *(i)* (Weak duality) *For every primal feasible X and dual feasible y:*

$$\langle C,\, X\rangle \;\leq\; b^\top y.$$

(ii) (Strong duality) *Under mild regularity conditions (e.g. strict feasibility – Slater's condition), the primal and dual optimal values are equal and both attained.*

Proof of weak duality. $b^\top y - \langle C,\, X\rangle = \sum_i y_i b_i - \langle C,\, X\rangle = \sum_i y_i \langle A_i,\, X\rangle - \langle C,\, X\rangle = \langle \sum_i y_i A_i - C,\, X\rangle \geq 0$, since $\sum_i y_i A_i - C \succeq 0$ and $X \succeq 0$ implies $\langle Z,\, X\rangle = \mathrm{tr}(ZX) \geq 0$ for any two PSD matrices Z, X. ∎

12.2.3 Computational complexity and algorithms

Theorem 12.5 (Polynomial-time solvability). *An SDP with $n \times n$ matrix variable and m constraints can be solved to additive accuracy ε in time $O(\mathrm{poly}(n, m) \log(1/\varepsilon))$.*

12.2.4 Vector programs: a useful reformulation

Many SDPs arising in combinatorics are most naturally written as *vector programs*: assign a vector $v_i \in \mathbb{R}^n$ to each object, subject to inner-product constraints.

Proposition 12.6 (Equivalence). *Specifying $X \in \mathcal{S}^n$ with $X \succeq 0$ is equivalent to specifying vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ with $X_{ij} = \langle v_i, v_j \rangle$ (a* Gram matrix representation*).*

Proof. $X \succeq 0 \Leftrightarrow X = V^\top V$ for some matrix V with columns $v_1, \ldots, v_n$. Then $X_{ij} = v_i^\top v_j = \langle v_i, v_j \rangle$. ∎

Diagonal constraints $X_{ii} = 1$ become $\|v_i\|^2 = 1$ (unit vectors).

12.3 MAX-CUT: The Problem

12.3.1 Definition

Definition 12.7 (MAX-CUT). Given an undirected graph $G = (V, E)$ with $|V| = n$, $|E| = m$, and non-negative edge weights $w : E \to \mathbb{R}_{\geq 0}$, the *MAX-CUT* problem asks for a partition $(S, \bar{S})$ of V that maximizes the total weight of edges crossing the cut:

$$\mathrm{OPT}(G) \; := \; \max_{S \subseteq V} \sum_{\{i,j\} \in E} w_{ij} \cdot \mathbf{1}[i \in S,\; j \notin S].$$

± 1 formulation. Label each vertex with $x_i \in \{+1, -1\}$ (convention: $x_i = +1$ for $i \in S$). An edge $\{i, j\}$ is cut iff $x_i \neq x_j$ iff $x_i x_j = -1$ iff $\frac{1 - x_i x_j}{2} = 1$. So:

$$\mathrm{OPT}(G) \; = \; \max_{x \in \{+1,-1\}^n} \frac{1}{2} \sum_{\{i,j\} \in E} w_{ij}(1 - x_i x_j). \tag{12.3}$$

12.4 The Goemans–Williamson SDP Relaxation

12.4.1 Formulation

Definition 12.8 (GW SDP relaxation – Gram matrix view). The *Goemans–Williamson SDP* for MAX-CUT is:

$$\mathrm{SDP}(G) \; := \; \max_{\{v_i\} \subset \mathbb{R}^n} \; \frac{1}{2} \sum_{\{i,j\} \in E} w_{ij}(1 - \langle v_i, v_j \rangle) \quad \text{s.t.} \quad \|v_i\|^2 = 1 \;\; \forall i \in V. \tag{12.4}$$

In matrix form, let $X = (X_{ij})$ with $X_{ij} = \langle v_i, v_j \rangle$. Then (12.4) becomes:

$$\max_{X \in \mathcal{S}^n} \quad \frac{1}{2} \sum_{\{i,j\}} w_{ij}(1 - X_{ij}) \quad \text{s.t.} \quad X_{ii} = 1 \; \forall i, \quad X \succeq 0.$$

Equivalently, with the Laplacian-like matrix $L_{ij} = -w_{ij}$ for $\{i, j\} \in E$ and $L_{ii} = \sum_j w_{ij}$:

$$\mathrm{SDP}(G) = \max_{X \succeq 0,\, \mathrm{diag}(X) = \mathbf{1}} \frac{1}{4} \langle L, X \rangle.$$

Proposition 12.9 (Relaxation). $\mathrm{OPT}(G) \le \mathrm{SDP}(G)$.

Proof. Any integer solution $x \in \{+1,-1\}^n$ gives a feasible SDP solution: set $v_i = x_i\mathbf{e}_1 \in \mathbb{R}^n$ (scalar multiples of the first standard basis vector). Then $\langle v_i, v_j\rangle = x_i x_j$, so the SDP objective equals the integer program objective. Since $\mathrm{SDP}(G)$ is a maximum over a *larger* feasible set, the inequality follows. ∎

12.4.2 Geometric picture

The SDP places one unit vector per vertex anywhere on the unit sphere $\mathbb{S}^{n-1}$. The objective wants to *spread the vectors apart*: an edge $\{i,j\}$ contributes $\frac{w_{ij}}{2}(1-\langle v_i, v_j\rangle)$, which is maximized (at $w_{ij}/2$) when v_i and v_j are *antipodal* ($\langle v_i, v_j\rangle = -1$, angle $= \pi$), and zero when they are aligned ($\langle v_i, v_j\rangle = 1$, angle $= 0$).

Key geometric intuition. The SDP solves a "spreading problem": embed the vertices as unit vectors on a sphere to maximize weighted pairwise separation (as measured by $1-\langle v_i, v_j\rangle$). Integer solutions correspond to the special case where all vectors are $\pm\mathbf{e}_1$. The SDP allows vectors to point in arbitrary directions, creating a potentially higher value.

12.5 The Hyperplane Rounding Algorithm

12.5.1 The algorithm

Given the SDP solution $\{v_i\}$, we must produce a cut $S \subseteq V$. The key idea is: a random hyperplane through the origin partitions the sphere into two hemispheres, and we assign each vertex to the hemisphere containing its vector.

Algorithm 12.10 (Goemans–Williamson hyperplane rounding).

Step 1. Solve the SDP. Compute unit vectors $v_1,\dots,v_n \in \mathbb{R}^n$ achieving (or ε-approximating) $\mathrm{SDP}(G)$.

Step 2. Sample a random hyperplane. Draw a random unit vector r uniformly from $\mathbb{S}^{n-1}$. (Equivalently, draw $r \sim \mathcal{N}(0, I_n)$ and normalize.)

Step 3. Round. For each vertex i, set

$$x_i \;:=\; \mathrm{sgn}\big(\langle r, v_i\rangle\big) \;\in\; \{+1,-1\}.$$

Step 4. Output. Return the cut $S = \{i \in V : x_i = +1\}$.

12.5.2 Why this works: the angle probability formula

The heart of the analysis is a clean geometric calculation.

Lemma 12.11 (Angle probability). *For any two unit vectors $u, v \in \mathbb{S}^{n-1}$ with $\langle u, v\rangle = \cos\theta$ (so $\theta \in [0,\pi]$ is the angle between them), and a uniformly random unit vector r:*

$$\mathbb{P}\big[\mathrm{sgn}(\langle r, u\rangle) \neq \mathrm{sgn}(\langle r, v\rangle)\big] \;=\; \frac{\theta}{\pi}. \tag{12.5}$$

Proof. The event $\text{sgn}(\langle r, u\rangle) \neq \text{sgn}(\langle r, v\rangle)$ occurs iff r lies in the region where $\langle r, u\rangle$ and $\langle r, v\rangle$ have opposite signs, i.e., r lies in one of the two "wedges" defined by the hyperplanes $\langle r, u\rangle = 0$ and $\langle r, v\rangle = 0$.

By symmetry, we may work in the 2-dimensional plane spanned by u and v. The two hyperplanes through the origin orthogonal to u and v intersect this plane in two lines, dividing the plane (and by rotation-invariance of r, the full sphere) into four wedge-shaped sectors. The two sectors where u and v have opposite signs each subtend an angle of θ at the origin (the angle between u and v), together covering a fraction $2\theta/(2\pi) = \theta/\pi$ of the circle. Since r is uniformly distributed, $\mathbb{P}[\text{opposite signs}] = \theta/\pi$. ∎

Remark 12.12. This formula is remarkable: the probability depends *only on the angle* θ, not on the dimension n or the specific directions.

12.6 Analysis: The Approximation Ratio

12.6.1 Expected cut value

Theorem 12.13 (Expected value of the rounded cut).

$$\mathbb{E}[\text{val}(S)] \;=\; \sum_{\{i,j\}\in E} w_{ij} \cdot \frac{\theta_{ij}}{\pi},$$

where $\theta_{ij} = \arccos(\langle v_i, v_j\rangle) \in [0, \pi]$ *is the angle between the SDP vectors for vertices* i *and* j*.*

Proof. By linearity of expectation:

$$\mathbb{E}[\text{val}(S)] = \sum_{\{i,j\}\in E} w_{ij} \cdot \mathbb{P}\big[\text{sgn}(\langle r, v_i\rangle) \neq \text{sgn}(\langle r, v_j\rangle)\big] = \sum_{\{i,j\}\in E} w_{ij} \cdot \frac{\theta_{ij}}{\pi},$$

using Lemma 12.11. ∎

12.6.2 The key analytical inequality

We now compare $\mathbb{E}[\text{val}(S)]$ to the SDP objective value.

Lemma 12.14 (GW inequality). *For all* $\theta \in [0, \pi]$*:*

$$\frac{\theta}{\pi} \;\geq\; \alpha_{\text{GW}} \cdot \frac{1-\cos\theta}{2}, \tag{12.6}$$

where

$$\alpha_{\text{GW}} \;:=\; \min_{\theta\in(0,\pi)} \frac{2\theta}{\pi(1-\cos\theta)} \;\approx\; 0.87856\ldots \tag{12.7}$$

The minimum is attained at $\theta^* \approx 2.3312$ *radians* ($\approx 133.6°$)*.*

Proof. Define $f(\theta) = \frac{2\theta}{\pi(1-\cos\theta)}$ on $(0, \pi)$. We verify:

1. $\lim_{\theta\to 0^+} f(\theta) = \frac{2}{\pi} \cdot \lim_{\theta\to 0} \frac{\theta}{1-\cos\theta} = \frac{2}{\pi} \cdot \lim_{\theta\to 0} \frac{1}{\sin\theta} = +\infty$ (L'Hôpital).

2. $f(\pi) = \frac{2\pi}{\pi \cdot 2} = 1$.

3. $f'(\theta) = 0$ has a unique solution θ^* in $(0, \pi)$ satisfying $\tan(\theta^*/2) = \theta^*/2 + \text{correction}$; numerically $\theta^* \approx 2.3312$.

4. $f(\theta^*) \approx 0.87856$.

So $f(\theta) \geq f(\theta^*) = \alpha_{\text{GW}}$ for all $\theta \in (0, \pi)$, and equality also holds trivially at the boundary $\theta = 0$ (both sides are 0) and $\theta = \pi$ (both sides give $1 \geq \alpha_{\text{GW}}$). ∎

12.6.3 The main theorem

Theorem 12.15 (Goemans–Williamson 1995). *Algorithm 12.10 is a polynomial-time randomized α_{GW}-approximation algorithm for MAX-CUT:*

$$\mathbb{E}[\text{val}(S)] \;\geq\; \alpha_{\text{GW}} \cdot \text{OPT}(G) \approx 0.8785 \cdot \text{OPT}(G).$$

Proof.

$$\begin{aligned}
\mathbb{E}[\text{val}(S)] &= \sum_{\{i,j\} \in E} w_{ij} \cdot \frac{\theta_{ij}}{\pi} && (12.8)\\
&\geq \alpha_{\text{GW}} \cdot \sum_{\{i,j\} \in E} w_{ij} \cdot \frac{1 - \cos\theta_{ij}}{2} && (12.9)\\
&= \alpha_{\text{GW}} \cdot \sum_{\{i,j\} \in E} w_{ij} \cdot \frac{1 - \langle v_i, v_j \rangle}{2} && (12.10)\\
&= \alpha_{\text{GW}} \cdot \text{SDP}(G) && (12.11)\\
&\geq \alpha_{\text{GW}} \cdot \text{OPT}(G). && (12.12)
\end{aligned}$$

Here (12.8) is Theorem 12.13, (12.9) applies Lemma 12.14 edge-by-edge, (12.10) uses $\cos\theta_{ij} = \langle v_i, v_j \rangle$, (12.11) is the SDP objective, and (12.12) uses the relaxation bound $\text{SDP}(G) \geq \text{OPT}(G)$. ∎

Summary of the proof in one line.
The GW analysis reduces to verifying a single analytical inequality:

$$\boxed{\frac{\theta}{\pi} \;\geq\; \alpha_{\text{GW}} \cdot \frac{1 - \cos\theta}{2} \qquad \forall\, \theta \in [0, \pi].}$$

The left side is the probability that a random hyperplane cuts the edge; the right side is α_{GW} times the SDP contribution of the edge.

12.7 PSD Matrices and Moments: The Bridge to SoS

12.7.1 The moment matrix

DEFINITION 12.16 (Moment matrix). Let μ be a probability distribution over $\mathbb{R}^n$ (or $\{0,1\}^n$, or $\{+1,-1\}^n$). Fix a degree $d \geq 1$, and let $[x]_d$ denote the vector of all monomials of degree $\leq d$:

$$[x]_d \;=\; \big(1,\; x_1,\; x_2,\; \ldots,\; x_n,\; x_1^2,\; x_1x_2,\; \ldots\big)^\top \;\in\; \mathbb{R}^{\binom{n+d}{d}}.$$

The *degree-d moment matrix* of μ is

$$M_d(\mu) \;:=\; \mathbb{E}_{x\sim\mu}\Big[[x]_d \cdot [x]_d^\top\Big].$$

Its (α,β)-entry is $M_d(\mu)_{\alpha\beta} = \mathbb{E}_\mu[x^\alpha x^\beta]$, the joint moment of monomials x^α and x^β.

EXAMPLE 12.17 (degree-2 moment matrix). For $d = 1$, $[x]_1 = (1, x_1, \ldots, x_n)^\top$, so

$$M_1(\mu) \;=\; \mathbb{E}\begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}\begin{pmatrix} 1 & x_1 & \cdots & x_n \end{pmatrix} \;=\; \begin{pmatrix} 1 & \mathbb{E}[x_1] & \cdots & \mathbb{E}[x_n] \\ \mathbb{E}[x_1] & \mathbb{E}[x_1^2] & \cdots & \mathbb{E}[x_1x_n] \\ \vdots & & \ddots & \vdots \\ \mathbb{E}[x_n] & \mathbb{E}[x_nx_1] & \cdots & \mathbb{E}[x_n^2] \end{pmatrix}.$$

The top-left 1 encodes normalization ($\mathbb{E}[1] = 1$); the off-diagonal entries are covariances (up to centering); the diagonal is the vector of second moments. For $x \in \{+1,-1\}^n$, we have $x_i^2 = 1$, so $\mathbb{E}[x_i^2] = 1$ and $M_1(\mu)_{ii} = 1$ for all $i \geq 1$.

12.7.2 The moment matrix is always PSD

THEOREM 12.18 (Moment matrices are PSD). *For any distribution μ and any degree d, $M_d(\mu) \succeq 0$.*

Proof. For any vector v of the same dimension as $[x]_d$:

$$v^\top M_d(\mu)\, v \;=\; v^\top \mathbb{E}\Big[[x]_d[x]_d^\top\Big]\, v \;=\; \mathbb{E}\Big[v^\top [x]_d[x]_d^\top v\Big] \;=\; \mathbb{E}\Big[\big(v^\top [x]_d\big)^2\Big] \;\geq\; 0.$$

The first equality is the definition of $M_d(\mu)$. The second uses linearity of expectation. The third rewrites $v^\top Av = (v^\top w)^2$ where $w = [x]_d$. The inequality holds because it is the expectation of a square. ∎

One-line summary. $v^\top M_d(\mu)\, v = \mathbb{E}\Big[\big(v^\top [x]_d\big)^2\Big] \geq 0$ because it is the *expected value of a squared quantity.* PSD-ness of the moment matrix is exactly *non-negativity of variance* in disguise.

12.7.3 The GW Gram matrix as a degree-2 moment matrix

The connection to the GW SDP is immediate.

For $x \in \{+1, -1\}^n$ distributed according to μ, the degree-2 moment matrix $M(\mu)$ (restricted to the x-block, i.e., ignoring the constant monomial row/column) has entries $M_{ij} = \mathbb{E}_\mu[x_i x_j]$ and diagonal entries $\mathbb{E}_\mu[x_i^2] = 1$.

This is *exactly* the Gram matrix X of the GW SDP, with $X_{ij} = \langle v_i, v_j \rangle$! More precisely:

PROPOSITION 12.19. *The GW SDP feasible set* $\{X \in \mathcal{S}^n : X \succeq 0, X_{ii} = 1\}$ *is precisely the set of degree-2 moment matrices of distributions over* $\{+1, -1\}^n$, relaxed *to all PSD matrices with unit diagonal.*

That is, every actual distribution μ over $\{+1, -1\}^n$ gives a valid GW SDP solution (its moment matrix), but the SDP also allows PSD matrices that do *not* come from any real distribution. This is the *integrality gap*: the SDP can "pretend" to be a distribution without actually being one.

The probabilistic perspective suggest a "different" rounding algorithm *Gaussian rounding*:

1. Sample $g \sim \mathcal{N}(0, X)$;
2. Set $x_i = \operatorname{sgn} g_i$.

A similar analysis as the random hyperplane rounding, using rotational invariance of multivariate Gaussian, gives the same α_{GW} approximation ratio.