Advanced Algorithms / Fall 2026

Linear Programming and Optimization

Semidefinite Programming

Matrix relaxations and Goemans–Williamson rounding

Lecture 9: 100 minutes

These notes cover only the linear-programming strand taught by Shi Li. This module comprises nine 100-minute lectures, numbered consistently with the slides. Parts I and II are reading divisions within this lecture; optional appendices are for further study.

Instructor: Shi Li

Nanjing University

Prepared from the instructor's course slides, with AI-assisted expansion of explanations, proofs, and independent-study exercises. The English and Chinese editions have matching mathematical content.

Learning goals and reading route

The final lecture enlarges the class of relaxations from linear programs to semidefinite programs. Part I develops the matrix language and the Max-Cut relaxation. Part II rounds its vectors with a random hyperplane and proves the approximation guarantee. SDP duality is provided as optional reading.

- Recognize equivalent quadratic, spectral, and Gram-matrix descriptions of positive semidefiniteness.
- Formulate Max-Cut as an SDP and separate an indefinite matrix using a negative-eigenvalue direction.
- Derive the angle-over-π cutting probability and the constant α_{GW}.

Prerequisites are elementary linear algebra, the spectral theorem for real symmetric matrices, expectation, and the separation-oracle viewpoint from Lecture 4. All edge weights below are nonnegative; the unweighted slides are the special case in which every edge has weight one.

Contents

1 Part I: From cuts to positive semidefinite matrices **1**
1.1 Max-Cut and the one-half baseline . . . 1
1.2 Why the naive linear relaxations are weak . . . 1
1.3 Replace signs by unit vectors . . . 1
1.4 Positive semidefinite matrices . . . 2
1.5 The matrix formulation and separation . . . 2

2 Part II: Goemans–Williamson random-hyperplane rounding **3**
2.1 The rounding algorithm . . . 3
2.2 The angle lemma . . . 3
2.3 The approximation constant . . . 3
2.4 A complete triangle example . . . 4
2.5 Precision, probability, and hardness . . . 4
2.6 Independent-study exercises . . . 5

A Optional: semidefinite programming duality **5**
A.1 The dual cone and weak duality . . . 5
A.2 Derivation and relation to LP . . . 6

1 Part I: From cuts to positive semidefinite matrices

1.1 Max-Cut and the one-half baseline

Given an undirected loopless graph $G = (V, E)$ with weights $w_e \geq 0$, Max-Cut asks for a partition $(S, V \setminus S)$ maximizing the total weight of crossing edges. Write $W = \sum_e w_e$. The complementary problem Min-Uncut minimizes the weight of edges whose endpoints lie on the same side. It is also the minimum weight of edges to remove to make the graph bipartite. Their exact optimum values sum to W, but a multiplicative approximation guarantee for one does not transfer with the same factor to the other.

Proposition 1.1. *There is a simple randomized* $1/2$*-approximation, and a greedy deterministic algorithm with the same guarantee.*

Proof. Independently place every vertex on each side with probability $1/2$. Each edge crosses with probability $1/2$, so the expected cut weight is $W/2 \geq \mathrm{OPT}/2$. For a greedy alternative, process vertices in any order and put the next vertex on the side that cuts at least half the total weight of its edges to previously processed vertices. Charge each edge when its second endpoint is placed; at least half the total weight is cut. □

Another proof uses local search: at a partition where moving any single vertex cannot improve the cut, each vertex has at least as much incident weight crossing as staying on its side. Summing over vertices yields cut weight at least $W/2$. For unweighted graphs, every improving move increases the integer cut size, so at most $|E|$ moves occur. This particular iteration bound does not extend to arbitrary binary-encoded weights.

1.2 Why the naive linear relaxations are weak

One tempting encoding uses side variables $y_v \in [0, 1]$ and cut variables x_{uv}, requiring $x_{uv} \leq |y_u - y_v|$. This is not a linear constraint or even a convex constraint: $(y_u, y_v, x_{uv}) = (1, 0, 1/2)$ and $(0, 1, 1/2)$ are feasible, while their midpoint $(1/2, 1/2, 1/2)$ is not. Replacing it by $x_{uv} \geq |y_u - y_v|$ would be linearizable, but it reverses the required relation. A valid LP uses $x_{uv} \in [0, 1]$ for every pair of vertices and the inequalities

$$x_{uv} + x_{vw} + x_{uw} \leq 2 \qquad (u, v, w \text{ distinct}).$$

A cut crosses at most two edges of any triangle, so every cut is feasible. However, on a triangle-free graph, assigning one to graph edges and zero to nonedges satisfies all these inequalities and has objective W. In the unweighted case there are triangle-free graphs whose maximum cut is arbitrarily close to $|E|/2$, so the relaxation's integrality gap approaches two. The existence statement can be proved by taking a sufficiently sparse random graph and deleting triangles; it is not true that every triangle-free graph has such a small maximum cut (bipartite graphs cut every edge).

1.3 Replace signs by unit vectors

Represent the side of vertex v by $s_v \in \{-1, 1\}$. Then $(1 - s_u s_v)/2$ is exactly the indicator that edge uv is cut. Hence

$$\mathrm{OPT} = \max_{s_v \in \{-1,1\}} \frac{1}{2} \sum_{uv \in E} w_{uv}(1 - s_u s_v).$$

Relax each sign to a unit vector $y_v \in \mathbb{R}^n$, where $n = |V|$:

$$\max \ \frac{1}{2} \sum_{uv \in E} w_{uv}(1 - \langle y_u, y_v \rangle) \quad \text{s.t. } \|y_v\|_2 = 1 \quad (v \in V). \tag{1.1}$$

Every cut embeds by choosing $y_v = s_v e_1$, so the optimum of this relaxation is at least OPT. The displayed vector formulation has quadratic expressions. Its convexity becomes visible after changing variables to the Gram matrix.

1.4 Positive semidefinite matrices

Definition 1.2. A real symmetric matrix X is positive semidefinite (PSD), written $X \succeq 0$, if $z^\mathsf{T} X z \geq 0$ for every real vector z. Write $X \succeq Y$ when $X - Y \succeq 0$. Positive definiteness, written $X \succ 0$, requires strict positivity for every nonzero z.

Theorem 1.3. *For a real symmetric $n \times n$ matrix X, the following are equivalent:*

1. $X \succeq 0$.
2. *Every eigenvalue of X is nonnegative.*
3. $X = V^\mathsf{T} V$ *for some real matrix V with at most n rows.*
4. $X = \sum_{k=1}^{n} \lambda_k q_k q_k^\mathsf{T}$, *where $\lambda_k \geq 0$ and the q_k form an orthonormal basis.*

Proof. The spectral theorem writes $X = Q \Lambda Q^\mathsf{T}$. Evaluating the quadratic form at each eigenvector shows that PSD implies $\lambda_k \geq 0$. Conversely, nonnegative eigenvalues give $z^\mathsf{T} X z = \sum_k \lambda_k (q_k^\mathsf{T} z)^2 \geq 0$. Taking $V = \Lambda^{1/2} Q^\mathsf{T}$ gives the Gram factorization; zero rows can be omitted. Finally, $X = V^\mathsf{T} V$ implies $z^\mathsf{T} X z = \|Vz\|_2^2 \geq 0$. The spectral sum is the same decomposition written term by term. □

The columns of V are vectors y_v with $X_{uv} = \langle y_u, y_v \rangle$. Thus any collection of n vectors in a higher-dimensional space has the same pairwise inner products as some vectors in $\mathbb{R}^n$. Positive semidefiniteness is not entrywise nonnegativity: $\left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$ is PSD. Nor do nonnegative entries imply PSD: $\left(\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right)$ has eigenvalue -1.

1.5 The matrix formulation and separation

For symmetric matrices use $A \bullet X = \operatorname{tr}(A^\mathsf{T} X) = \sum_{u,v} A_{uv} X_{uv}$. The off-diagonal entries are counted twice in this full-matrix inner product. A general SDP has a linear matrix objective, linear scalar constraints, and a PSD constraint, for example

$$\min C \bullet X \quad \text{s.t. } A_k \bullet X \geq b_k \ (k \in [m]), \quad X \succeq 0.$$

The numerical example in the slides uses $X = \left(\begin{smallmatrix} z_1 & z_2 \\ z_2 & z_3 \end{smallmatrix}\right)$, objective $5z_1 + 6z_2 + 7z_3$, and constraints $z_1 + 3z_2 + 4z_3 \geq 5$, $2z_1 + 3z_2 + z_3 \geq 10$, and $3z_1 + 2z_2 + 2z_3 \geq 7$. For example, its objective matrix is $C = \left(\begin{smallmatrix} 5 & 3 \\ 3 & 7 \end{smallmatrix}\right)$, not a matrix with off-diagonal entries six.

The Max-Cut SDP

$$\mathrm{SDP} = \max \frac{1}{2} \sum_{uv \in E} w_{uv}(1 - X_{uv}) \quad \text{s.t. } X_{vv} = 1 \ (v \in V), \quad X \succeq 0. \tag{1.2}$$

The PSD cone is convex because every inequality $z^\mathsf{T} X z \geq 0$ is linear in X for fixed z. The feasible set in (1.2) is compact: its entries satisfy $|X_{uv}| \leq \sqrt{X_{uu} X_{vv}} = 1$.

To separate a symmetric matrix X from the PSD cone, find an eigenvector z with negative eigenvalue. Then $z^\mathsf{T} X z < 0$, whereas every PSD matrix Y satisfies $(zz^\mathsf{T}) \bullet Y \geq 0$. This is a separating linear inequality in the matrix entries. Check the other linear constraints directly. This explains the ellipsoid connection, but numerical precision and conditioning still matter. General SDPs do not inherit all exact feasibility and attainment properties of rational LPs. For Max-Cut, I is a strictly

feasible matrix relative to the diagonal equalities, and the feasible set is bounded. Standard SDP algorithms compute an approximately optimal feasible solution to prescribed accuracy. In the analysis below we first use an ideal optimum and then state the effect of objective error.

2 Part II: Goemans–Williamson random-hyperplane rounding

2.1 The rounding algorithm

Factor a feasible SDP solution into unit vectors y_v, and use one common random direction for every vertex. Sharing the direction preserves information about pairwise angles. Independently choosing a random side for each vertex would discard that information.

Random-hyperplane rounding

1. Sample $g \in \mathbb{R}^n$ with independent coordinates $g_k \sim N(0,1)$.
2. Set $S = \{v : \langle g, y_v \rangle \geq 0\}$ and return $(S, V \setminus S)$.

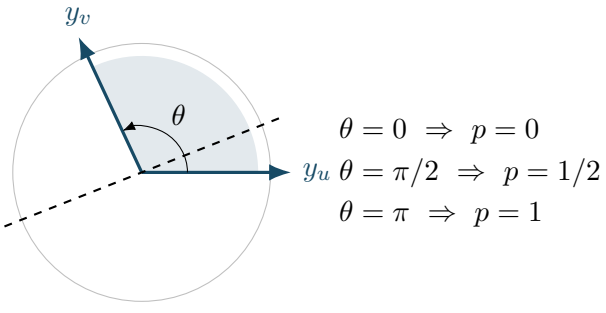
The boundary event has probability zero for any fixed unit vector, so the tie convention does not affect the analysis. Normalizing g is unnecessary because only the signs of its inner products are used.

2.2 The angle lemma

Lemma 2.1. *If unit vectors y_u, y_v have angle $\theta \in [0, \pi]$, then*

$$\mathbb{P}[uv \text{ is cut}] = \frac{\theta}{\pi} = \frac{\arccos\langle y_u, y_v \rangle}{\pi}.$$

Proof. First assume the vectors are linearly independent and project g onto their two-dimensional span. The signs of the two inner products depend only on this projection. Rotational symmetry of the Gaussian makes its direction uniform on the circle. The directions with nonnegative inner product with each vector form a semicircle. The symmetric difference of these two semicircles has total angular measure 2θ, giving probability $2\theta/(2\pi)$. If the vectors agree, the probability is zero; if they are opposite, it is one. These are the remaining cases. □



2.3 The approximation constant

For one edge, write $t = \langle y_u, y_v \rangle = \cos\theta$. Its SDP contribution, before multiplying by its weight, is $(1-t)/2$. Its expected rounded contribution is $\arccos(t)/\pi$. Define

$$\alpha_{\mathrm{GW}} = \min_{-1 \leq t < 1} \frac{2\arccos t}{\pi(1-t)} = \min_{0 < \theta \leq \pi} \frac{2\theta}{\pi(1-\cos\theta)} \approx 0.878567.$$

The ratio diverges as $\theta \to 0$ and equals one at $\theta = \pi$. Its derivative has the sign of

$$h(\theta) = 1 - \cos\theta - \theta \sin\theta, \qquad h'(\theta) = -\theta \cos\theta.$$

Thus h decreases from zero to a negative value on $(0, \pi/2)$ and then increases strictly to 2 on $(\pi/2, \pi)$. It has exactly one zero in the latter interval, which gives the unique minimum. Equivalently, the minimizing angle satisfies $\tan(\theta/2) = \theta$; it is approximately 2.33112 radians. The definition by a minimum is exact, while the displayed decimal is a numerical evaluation.

Theorem 2.2 (Goemans–Williamson guarantee)**.** *Random-hyperplane rounding of an optimal Max-Cut SDP solution produces a cut C satisfying*

$$\mathbb{E}[w(C)] \geq \alpha_{\mathrm{GW}}\,\mathrm{SDP} \geq \alpha_{\mathrm{GW}}\ \mathrm{OPT}\,.$$

Proof. For each edge with $t < 1$, the definition of α_{GW} gives $\arccos(t)/\pi \geq \alpha_{\mathrm{GW}}(1-t)/2$. If $t = 1$, both sides are zero. Multiply by $w_{uv} \geq 0$ and sum. Linearity of expectation does not require the edge-cut events to be independent:

$$\begin{aligned}\mathbb{E}[w(C)] &= \sum_{uv \in E} w_{uv}\mathbb{P}[uv \text{ is cut}] \\ &\geq \alpha_{\mathrm{GW}} \frac{1}{2} \sum_{uv \in E} w_{uv}(1 - \langle y_u, y_v\rangle) \\ &= \alpha_{\mathrm{GW}}\,\mathrm{SDP} \geq \alpha_{\mathrm{GW}}\ \mathrm{OPT}\,.\end{aligned}$$

□

2.4 A complete triangle example

For an unweighted triangle, every nontrivial cut has value two. Place its three unit vectors at angles 0, $2\pi/3$, and $4\pi/3$. Every pair has inner product $-1/2$, so the SDP objective is $3 \cdot (3/4) = 9/4$. This is optimal, because for any three unit vectors

$$\frac{1}{2}\sum_{u<v}(1 - \langle y_u, y_v\rangle) = \frac{9 - \|y_1 + y_2 + y_3\|_2^2}{4} \leq \frac{9}{4}.$$

Rounding cuts each edge with probability $2/3$, so the expected cut value is exactly two. In fact, almost every hyperplane produces a nontrivial cut. This instance has ratio $2/(9/4) = 8/9$ between rounded value and SDP value, slightly above the universal constant.

2.5 Precision, probability, and hardness

The theorem is an expectation guarantee, not a claim that every sampled hyperplane attains the factor. Repeating independently and keeping the best cut cannot decrease expected quality. More quantitatively, for $\mathrm{OPT} > 0$ and $0 < \varepsilon < \alpha_{\mathrm{GW}}$, the bound $0 \leq w(C)/\,\mathrm{OPT} \leq 1$ implies

$$\mathbb{P}[w(C) \geq (\alpha_{\mathrm{GW}} - \varepsilon)\,\mathrm{OPT}] \geq \frac{\varepsilon}{1 - \alpha_{\mathrm{GW}} + \varepsilon}.$$

To see this, if the success probability is p, the expectation of the normalized value is at most $p + (1-p)(\alpha_{\mathrm{GW}} - \varepsilon)$. Rearranging gives the stated bound. Repetition amplifies the success probability. If $\mathrm{OPT} = 0$, every cut is optimal. If the computed feasible SDP solution has objective at least $\mathrm{SDP} - \delta$, the same proof gives $\mathbb{E}[w(C)] \geq \alpha_{\mathrm{GW}}(\mathrm{OPT} - \delta)$. For $W > 0$, the baseline $\mathrm{OPT} \geq W/2$ permits choosing an accuracy proportional to W to obtain any desired additional relative error. The feasible Gram matrix and its normalization must be handled consistently when

implementing a numerical solver. The original algorithm is due to Goemans and Williamson (1995). The slides also mention a conditional optimality result: assuming the Unique Games Conjecture, achieving any fixed factor above α_{GW} is NP-hard, as established by Khot, Kindler, Mossel, and O'Donnell. The resulting exclusion of polynomial-time algorithms also uses the usual $P \neq NP$ assumption; this is not an unconditional optimality theorem.

2.6 Independent-study exercises

1. Give numerical values of W, Max-Cut optimum, and a cut value showing why a constant Max-Cut ratio gives no constant Min-Uncut ratio.
2. Show that a symmetric 2×2 matrix $\left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ is PSD exactly when $a, c \geq 0$ and $ac - b^2 \geq 0$.
3. Find a separating inequality for $X = \left(\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right)$ using its negative eigenvalue.
4. Write the three constraint matrices for the numerical SDP example. Check the factors of two on their off-diagonal entries.
5. Verify the cutting probabilities for identical, orthogonal, and opposite vectors. Explain why edge-cut events need not be independent.
6. Derive the triangle SDP optimum using the Gram matrix instead of the vector identity. Then find a matching dual upper-bound certificate.
7. Using the probability bound above, how many independent rounds suffice for failure probability at most $\beta \in (0, 1)$ at target ratio $\alpha_{\mathrm{GW}} - \varepsilon$?

A Optional: semidefinite programming duality

A.1 The dual cone and weak duality

Lemma A.1. *For symmetric matrices $X, Z \succeq 0$, $X \bullet Z \geq 0$. Conversely, if a symmetric Z has nonnegative inner product with every PSD matrix, then $Z \succeq 0$.*

Proof. By cyclicity of trace, $\operatorname{tr}(XZ) = \operatorname{tr}(Z^{1/2} X Z^{1/2}) \geq 0$, because the matrix inside the final trace is PSD. For the converse, test against $X = zz^{\mathsf{T}}$ for every vector z, obtaining $z^{\mathsf{T}} Z z \geq 0$. □

Consider the following primal and dual programs:

$$\begin{aligned} \text{Primal:} \quad & \min C \bullet X \\ & A_k \bullet X \geq b_k \quad (k \in [m]), \quad X \succeq 0; \\ \text{Dual:} \quad & \max b^{\mathsf{T}} y \\ & y \geq 0, \quad Z := C - \textstyle\sum_k y_k A_k \succeq 0. \end{aligned}$$

For feasible X, y, their objective gap is

$$C \bullet X - b^{\mathsf{T}} y = Z \bullet X + \sum_k y_k (A_k \bullet X - b_k) \geq 0.$$

This proves weak duality directly. It also gives complementary slackness when equality of objectives is achieved: $Z \bullet X = 0$ and $y_k (A_k \bullet X - b_k) = 0$ for every k.

A.2 Derivation and relation to LP

The Lagrangian for the inequalities is

$$\mathcal{L}(X, y) = C \bullet X + \sum_k y_k(b_k - A_k \bullet X) = b^\mathsf{T} y + Z \bullet X, \qquad y \geq 0.$$

Its infimum over $X \succeq 0$ equals $b^\mathsf{T} y$ if $Z \succeq 0$. If Z has a negative direction z, taking $X = tzz^\mathsf{T}$ and $t \to \infty$ makes the infimum $-\infty$. Maximizing this lower bound gives the displayed dual. This finite-dimensional derivation avoids treating an infinite list of PSD inequalities as an ordinary finite LP. The slides give the same dual through the representation of a PSD matrix as a nonnegative sum of rank-one matrices zz^T. Such a representation needs at most n terms by the spectral theorem; no genuinely infinite sum is required. If all data are diagonal, $X \succeq 0$ implies nonnegative diagonal entries, and replacing X by its diagonal preserves all objectives and constraints. The primal and dual then reduce to the familiar LP pair $\min\{c^\mathsf{T} x : Ax \geq b, x \geq 0\}$ and $\max\{b^\mathsf{T} y : A^\mathsf{T} y \leq c, y \geq 0\}$.

Remark A.2 (Strong duality needs a condition). Unlike finite LP, general SDP can have a duality gap or nonattainment without regularity assumptions. A sufficient Slater condition for the displayed inequality-form primal is the existence of $X \succ 0$ with every $A_k \bullet X > b_k$. If its optimal value is finite, this condition gives zero duality gap and attainment of the dual optimum. Equality constraints are kept as equalities and require the corresponding relative-interior version; their dual multipliers are unrestricted.

For Max-Cut, let L be the weighted graph Laplacian. The objective equals $L \bullet X/4$, so its dual is

$$\min \sum_v z_v \quad \text{s.t. } \operatorname{Diag}(z) - L/4 \succeq 0, \qquad z \in \mathbb{R}^n.$$

The primal matrix I is strictly feasible relative to the diagonal equalities, and sufficiently large equal z_v are strictly dual feasible. Thus this particular SDP has the strong duality properties needed for matching numerical upper and lower certificates. On the unweighted triangle, $z_v = 3/4$ for all v gives dual value $9/4$, matching the vector construction.

Source and reading notes

Primary source: the instructor's slide folder

`9.SDP`

`Topics.tex` and `Optional.tex`.

Optional sections follow the optional slides where present. Explanatory examples and exercises are added to support independent reading; they do not change the lecture allocation.

Course website. The early-course notes linked there were used as examples of organization and level of detail, not as the source of this module's subject matter.