Advanced Algorithms / Fall 2026

Linear Programming and Optimization

Extension Complexity

Extended formulations and nonnegative rank

Lecture 8: 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

An exponential number of facets need not mean that every LP model is large: extra variables can make a description much smaller. Part I constructs an extended formulation of the permutahedron. Part II proves that the smallest possible extension is characterized exactly by a matrix factorization parameter.

- Distinguish a polytope, an extension, and the projection map.
- Verify both inclusions in a projection proof and count inequalities rather than merely variables.
- Use LP duality to turn an extended formulation into a nonnegative factorization of a slack matrix.

Core prerequisites are the bipartite matching polytope and LP duality. The spanning-tree formulation is optional, matching the slide appendix; it also uses the spanning-tree polytope and flow concepts. The final lower-bound results are statements for perspective; their research proofs are outside this lecture.

Contents

1 Part I: Smaller LPs in a larger space **1**

1.1 Extensions and what their size measures 1

1.2 The equality-plus-nonnegativity form 1

1.3 The permutahedron 1

2 Part II: Slack matrices and nonnegative rank **2**

2.1 A factorization without cancellation 2

2.2 The slack matrix of a polytope 2

2.3 The slack-matrix theorem 3

2.4 Additional rows, columns, and lower bounds 4

2.5 Independent-study exercises 4

A Optional: a compact formulation of spanning trees **4**

A.1 Optional exercise 5

1 Part I: Smaller LPs in a larger space

1.1 Extensions and what their size measures

For a family of combinatorial solutions $\mathcal{S} \subseteq 2^{[n]}$, let $P = \text{conv}\{\chi^S : S \in \mathcal{S}\}$. Optimizing a linear objective over P solves the combinatorial optimization problem exactly. Within its affine hull, a full irredundant inequality description has one inequality per facet. The number of facets can be exponential.

Definition 1.1. An extension of a nonempty polytope $P \subseteq \mathbb{R}^n$ is a polyhedron $Q \subseteq \mathbb{R}^{n+r}$ such that

$$P = \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^r,\ (x, y) \in Q\}.$$

Its size is the number of inequalities in an irredundant description within its affine hull; affine equations do not count. The extension complexity $\text{xc}(P)$ is the minimum size of any extension.

The objective stays $c^\mathsf{T} x$: auxiliary variables do not receive arbitrary new costs. A small rational extended formulation supports efficient LP optimization when its coefficients, equations, and variables also have a suitably bounded encoding. Extension complexity is primarily a geometric inequality count; it does not by itself bound coefficient bit lengths or provide a construction algorithm.

1.2 The equality-plus-nonnegativity form

Proposition 1.2. *Equivalently,* $\text{xc}(P)$ *is the smallest* r *for which matrices* E, F *and a vector* g *satisfy*

$$P = \{x : \exists y \in \mathbb{R}^r_{\geq 0},\ Ex + Fy = g\}.$$

Proof. Such a system defines an extension with at most r facets, because its only inequalities are the r nonnegativity constraints. Conversely, take an extension with r facet inequalities and convert each inequality into an equality using one nonnegative slack variable. There may still be unrestricted auxiliary variables z. The resulting equations have the form

$$Bz = h - Dx - Hy, \qquad y \geq 0.$$

Let the rows of N be a basis of the left nullspace of B. A vector lies in the column space of B exactly when N annihilates it. Eliminating z therefore leaves only $N(Dx + Hy) = Nh$ and $y \geq 0$, without adding inequalities. □

1.3 The permutahedron

Let P_n be the convex hull of all permutations of $(1, 2, \ldots, n)$, with $n \geq 2$. Every point satisfies $\sum_i x_i = n(n+1)/2$. Differences obtained by swapping the value in one fixed position with each other position span the zero-sum subspace, so $\dim P_n = n - 1$.

Proposition 1.3. *Every nonempty proper subset* $S \subsetneq [n]$ *defines a facet inequality*

$$\sum_{i \in S} x_i \geq \frac{|S|(|S|+1)}{2}.$$

Proof. If $|S| = k$, the smallest possible sum in these positions is $1 + \cdots + k$, so the inequality is valid. Equality holds precisely for permutations that place $1, \ldots, k$ in S and $k+1, \ldots, n$ in its complement. Swaps within S span a $(k-1)$-dimensional space; swaps within its complement span an $(n-k-1)$-dimensional space. These spaces have disjoint supports and hence independent

directions. The equality face has dimension at least $n-2$, and the extra equality bounds it above by $n-2$. Thus it is a facet. □

These are 2^n-2 distinct facets. Yet a permutation can also be represented by a permutation matrix Y, with $Y_{ij}=1$ exactly when position i receives value j. The projection is linear: $x_i=\sum_j jY_{ij}$. For example,

$$(3,1,2) \quad \longleftrightarrow \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Theorem 1.4 (A quadratic-size extension)**.**

$$\sum_i Y_{ij}=1 \quad (j\in[n]), \qquad \sum_j Y_{ij}=1 \qquad (i\in[n]),$$
$$Y_{ij}\geq 0 \quad (i,j\in[n]), \qquad x_i=\sum_j jY_{ij} \quad (i\in[n])$$

is an extended formulation of P_n, so $\mathrm{xc}(P_n)\leq n^2$.

Proof. The first three conditions describe the doubly stochastic matrices. This is the perfect matching polytope of the complete bipartite graph, so by Lecture 2 every feasible Y is a convex combination of permutation matrices. Applying the linear projection gives a convex combination of permutation vectors. Conversely, every permutation vector has its permutation-matrix lift, and convex combinations of those lifts remain feasible. The only inequalities are the n^2 nonnegativity constraints. □

2 Part II: Slack matrices and nonnegative rank

2.1 A factorization without cancellation

Definition 2.1. For $M\in\mathbb{R}^{a\times b}_{\geq 0}$, the nonnegative rank $\mathrm{rank}_+(M)$ is the smallest $r\geq 0$ such that $M=LU$ with $L\in\mathbb{R}^{a\times r}_{\geq 0}$ and $U\in\mathbb{R}^{r\times b}_{\geq 0}$. Equivalently, M is a sum of r nonnegative rank-one matrices.

$$\mathrm{rank}(M)\leq \mathrm{rank}_+(M)\leq \min\{a,b\}.$$

The lower bound follows from the rank of a product. For the upper bound, use either $M=I_aM$ or $M=MI_b$. Ordinary rank allows positive and negative terms to cancel; nonnegative rank does not. In data-analysis applications, columns may represent documents and rows words: a nonnegative factorization represents each document as a nonnegative mixture of latent topics. The slides' movie example has the analogous latent-feature interpretation. Exact nonnegative rank is NP-hard to compute; this is a separate theorem, not a consequence of the elementary inequalities above (see Vavasis).

2.2 The slack matrix of a polytope

Let $P=\{x : Ax\leq b,\ Hx=h\}$ be a nonempty positive-dimensional polytope. The equations describe its affine hull, and the rows of A correspond to its facets. List its vertices as $v^1,\ldots,v^N$.

Definition 2.2. The slack matrix is $S\in\mathbb{R}^{m\times N}_{\geq 0}$ with

$$S_{ij}=b_i-a_i^{\mathsf{T}}v^j.$$

A zero records incidence: vertex v^j lies on facet i. A positive slack measures how far the inequality is from equality; it equals Euclidean distance to the hyperplane only after normalizing $\|a_i\|_2 = 1$. Positive row scaling changes the numbers but not the nonnegative rank.

Example 2.3 (A triangle)**.** For $P = \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}$ with vertices $(0,0), (1,0), (0,1)$,

$$S = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \operatorname{rank}(S) = \operatorname{rank}_+(S) = 3.$$

The nonnegative rank is at least three by ordinary rank and at most three by the number of rows.

2.3 The slack-matrix theorem

Theorem 2.4 (Yannakakis)**.** *For the preceding polytope and slack matrix,*

$$\boxed{\operatorname{xc}(P) = \operatorname{rank}_+(S).}$$

From a factorization to an extension. Suppose $S = LU$ is a nonnegative factorization with inner dimension r. Consider

$$Q = \{(x, z) : Hx = h,\ Ax + Lz = b,\ z \geq 0\}.$$

Its projection lies in P, because $Lz \geq 0$ implies $Ax \leq b$. Conversely, the jth vertex lifts using $z = U_{\cdot j}$, since $Av^j + LU_{\cdot j} = b$. Every point of P is a convex combination of vertices, and the same combination of their lifts belongs to Q. Hence Q projects exactly to P and has only r inequalities, proving $\operatorname{xc}(P) \leq \operatorname{rank}_+(S)$. □

From an extension to a factorization. Use the equivalent form $P = \{x : \exists z \geq 0,\ Ex + Fz = g\}$ with r coordinates in z. For each vertex v^j, fix one feasible lift z^j. For facet i, consider the LP

$$\max\ a_i^{\mathsf{T}} x \quad \text{s.t. } Ex + Fz = g,\ z \geq 0, \qquad x \text{ free.}$$

Its optimum is exactly b_i: the facet inequality is valid and tight somewhere in P. LP duality supplies an unrestricted vector μ^i such that

$$E^{\mathsf{T}}\mu^i = a_i, \qquad F^{\mathsf{T}}\mu^i \geq 0, \qquad g^{\mathsf{T}}\mu^i = b_i.$$

Therefore

$$S_{ij} = b_i - a_i^{\mathsf{T}} v^j = (\mu^i)^{\mathsf{T}}(g - Ev^j) = (\mu^i)^{\mathsf{T}} F z^j.$$

Define row $L_{i\cdot} = (\mu^i)^{\mathsf{T}} F \geq 0$ and column $U_{\cdot j} = z^j \geq 0$. Then $S = LU$ with inner dimension r, proving $\operatorname{rank}_+(S) \leq \operatorname{xc}(P)$. □

Where the assumptions matter

Keep the affine-hull equations in the constructed extension; omitting them can enlarge the projection. In the reverse proof, tightness of the supporting inequality makes the dual optimum exactly b_i, eliminating a possible extra constant slack. For a one-point polytope, equations alone suffice and the extension complexity is zero; it can be treated separately.

2.4 Additional rows, columns, and lower bounds

Adding the slack column of any nonvertex point of P does not increase nonnegative rank: it is the same convex combination of vertex columns. Similarly, adding the row of a valid supporting inequality that is tight on a nonempty face does not increase the rank. By LP duality, its slack on P is a nonnegative combination of facet slacks, with affine-hull equations contributing zero. These claims concern supporting inequalities; a generic redundant inequality with a positive constant offset requires separate justification. The theorem gives a lower-bound strategy: find a slack submatrix whose nonnegative rank is large. The celebrated results mentioned in the slides apply to the following families.

- The traveling-salesman polytope is the convex hull of Hamiltonian-cycle incidence vectors of K_n.
- The cut polytope is $\operatorname{conv}\{\chi^{\delta(S)} : S \subseteq V\}$, with cuts in a complete graph. Here these are cut boundaries, unlike the upward-closed deletion cuts in Lecture 4.
- The correlation polytope is $\operatorname{conv}\{bb^{\mathsf{T}} : b \in \{0,1\}^n\}$.

Fiorini, Massar, Pokutta, Tiwary, and de Wolf established exponential lower bounds for these families using connections with communication complexity; see their paper. Rothvoss proved $2^{\Omega(n)}$ extension complexity for the perfect matching polytope of the complete graph on an even number n of vertices; see the matching-polytope paper. The general matching polytope inherits this lower bound because its maximum-cardinality face is the perfect matching polytope. The matching result makes a crucial distinction: polynomial-time optimization and separation do not imply a polynomial-size extended formulation. The lower bound rules out a small exact LP representation; it does not rule out polynomial-time matching algorithms. These lower bounds are unconditional geometric results, not deductions from $P \neq NP$.

2.5 Independent-study exercises

1. Write all six vertices of P_3 and identify the six facet inequalities. Construct a fractional doubly stochastic lift of $(2, 2, 2)$.
2. Write the slack matrix of the square $[0, 1]^2$. Show that its ordinary rank is three. Can you show that its nonnegative rank is four?
3. Prove invariance of nonnegative rank under positive row and column scaling and under row and column permutations.
4. Prove that a face of a polytope has extension complexity at most that of the polytope: add the exposing equality to an extension.
5. Given a factorization of a slack matrix, explicitly lift a point written as a convex combination of vertices. Why can a small ordinary-rank factorization fail to give an extension?
6. In a nonnegative rank-one matrix, the positive entries form a rectangle of row and column indices. Use this to obtain a support-based lower bound on nonnegative rank.

A Optional: a compact formulation of spanning trees

Let $G = (V, E)$ be a connected loopless undirected graph with $n \geq 2$ vertices. Recall

$$P_{\mathrm{ST}} = \{x \geq 0 : x(E) = n - 1, \quad x(E[S]) \leq |S| - 1 \quad (\emptyset \neq S \subsetneq V)\}.$$

Choose a root r and replace each undirected edge by two directed arcs. Use y_{uv} for fractional orientation and a separate flow f^v for each target $v \neq r$. Let $\delta^+(u)$ and $\delta^-(u)$ denote outgoing and incoming arcs.

The extended spanning-tree system

$$
\begin{aligned}
y(\delta^-(u)) &= 1 && (u \neq r),\\
y(\delta^-(r)) &= 0, && y \geq 0,\\
x_{\{u,v\}} &= y_{uv} + y_{vu} && (\{u,v\} \in E),\\
0 \leq f^v_a &\leq y_a && (a \text{ an arc},\ v \neq r),\\
f^v(\delta^+(u)) - f^v(\delta^-(u)) &= \begin{cases} 1 & u = r,\\ -1 & u = v,\\ 0 & \text{otherwise} \end{cases} && (u \in V,\ v \neq r).
\end{aligned}
$$

The flows share the same capacity vector y, but each flow separately may use the full capacities. There is no constraint $\sum_v f^v_a \leq y_a$. Imposing that shared-capacity constraint would incorrectly exclude many trees.

Theorem A.1. *The projection of this system onto x is exactly P_{ST}. Its size is $O(n|E|)$, and therefore $O(n^3)$ for the complete graph.*

Proof. Orient any tree away from r, and send one unit to v on its unique root-to-v path. This lifts every tree incidence vector, hence their entire convex hull. Conversely, a feasible lift has total arc weight $n-1$, by summing all indegree equations, so $x(E) = n-1$. Fix a nonempty proper set S.

- If $r \in S$, the total weight of arcs with head in S is $|S| - 1$. Internal arcs form a subset, so $x(E[S]) \leq |S| - 1$.
- If $r \notin S$, choose $v \in S$. The unit r–v flow forces $y(\delta^-(S)) \geq 1$. Summing indegrees over S gives $x(E[S]) + y(\delta^-(S)) = |S|$, proving the same inequality.

Thus the projection satisfies the spanning-tree description. There are $O(n|E|)$ flow nonnegativity and capacity inequalities. □

The slides contrast this cubic upper bound on K_n with a quadratic lower bound. The construction above establishes the upper bound; it does not identify the optimal extension size.

A.1 Optional exercise

1. Give a rooted tree that becomes infeasible if the spanning-tree extension mistakenly imposes $\sum_v f^v_a \leq y_a$.

Source and reading notes

Primary source: the instructor's slide folder

`8.ExtensionComplexity`

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

Course website. The spanning-tree appendix is optional; exercises are for independent study.