

Linear Programming, Flow Problems, and Discrete Optimal Transport

Edoardo Annunziata

Tommaso Scardecchia

Benedetta Selmi

Ernest Andrei Borbely

Contents

1	Linear programming foundations	2
1.1	Standard form and geometry	2
1.2	Bases, basic feasible solutions, and reduced costs	2
1.3	Fundamental simplex theorems	2
1.4	The simplex algorithm	5
2	Duality and complementary slackness	6
2.1	The dual linear program	6
2.2	Weak duality	7
2.3	Geometric prerequisites: projection and separation	7
2.4	Strong duality	8
2.5	Complementary slackness	9
2.6	Duality for general linear programs	10
2.7	A worked example	10
2.8	Outlook: from duality to flows and discrete OT	11
3	Network flows as linear programs	12
4	The discrete Kantorovich problem as an LP	12
4.1	Dual formulation	13
5	Birkhoff-von Neumann theorem	13
6	Algorithms for discrete optimal transport	15
6.1	Hungarian algorithm	15
6.2	Dantzig's transportation simplex	15
7	Conclusion	16

Goal of the notes. These notes connect linear programming theory with concrete algorithms for discrete optimal transport. We first recall the geometric foundations of linear programming and the simplex method. We then develop the primal-dual viewpoint in detail: weak duality explains why dual feasible points certify bounds, strong duality explains why the best certificate is exact, and complementary slackness gives the algebraic bridge between optimal primal and dual solutions. This bridge is then used for network flows, the discrete Kantorovich problem, the Birkhoff-von Neumann theorem, and classical algorithms such as the Hungarian method and Dantzig's transportation simplex.

1 Linear programming foundations

1.1 Standard form and geometry

We work with the standard-form minimization problem

$$\min\{c^\top x : Ax = b, x \geq 0\},$$

where $A \in \mathbb{R}^{m \times n}$ has rank m , $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and $x \in \mathbb{R}^n$ is the vector of decision variables. The feasible set is

$$P := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}.$$

Geometrically, P is a feasible polyhedron: the intersection of the affine space $Ax = b$ with the nonnegativity half-spaces. For $x \in P$ we write

$$I(x) := \{i : x_i = 0\},$$

namely the set of active nonnegativity constraints.

1.2 Bases, basic feasible solutions, and reduced costs

Let $B \subseteq \{1, \dots, n\}$ with $|B| = m$ and assume that the submatrix A_B is invertible. Then B is called a basis. If $N := \{1, \dots, n\} \setminus B$, the basic solution associated with B is obtained by setting

$$x_N = 0, \quad x_B = A_B^{-1}b.$$

If $x_B \geq 0$, then the solution is feasible and is called a basic feasible solution (BFS).

For each $j \in N$, the reduced cost is

$$\bar{c}_j := c_j - c_B^\top A_B^{-1} A_j.$$

It measures the marginal variation of the objective when the nonbasic variable x_j is allowed to enter the basis while the equality constraint $Ax = b$ is preserved.

1.3 Fundamental simplex theorems

The following statements are the geometric core of the simplex method.

Definition 1.1 (Vertex). A point $x \in P$ is a vertex of P if it cannot be written as a nontrivial convex combination of two distinct points of P .

Theorem 1.2 (Vertex $\iff$ BFS). For $x \in P$, the following are equivalent:

- (i) x is a vertex of P ;
- (ii) x is a basic feasible solution.

Proof. Suppose first that x is the BFS associated with a basis B , and assume

$$x = \lambda u + (1 - \lambda)v, \quad u, v \in P, \quad \lambda \in (0, 1).$$

For every $j \notin B$ we have $x_j = 0$. Since $u, v \geq 0$, necessarily $u_j = v_j = 0$ for every $j \notin B$. Moreover, $Au = Av = b$, hence

$$A_B u_B = b = A_B v_B.$$

Since A_B is invertible, $u_B = v_B = x_B$, so $u = v = x$. Thus x is a vertex.

Conversely, let x be a vertex and define $J := \{j : x_j > 0\}$. If the columns $\{A_j\}_{j \in J}$ were linearly dependent, there would exist a vector $u \neq 0$ such that

$$Au = 0, \quad u_j = 0 \quad \text{for } j \notin J.$$

For sufficiently small $\varepsilon > 0$, both $x + \varepsilon u$ and $x - \varepsilon u$ remain feasible: the equality constraint is preserved because $Au = 0$, and nonnegativity is preserved because only strictly positive components of x are perturbed. Hence

$$x = \frac{1}{2}(x + \varepsilon u) + \frac{1}{2}(x - \varepsilon u),$$

contradicting the fact that x is a vertex. Therefore the columns indexed by J are linearly independent. Completing them to a basis B , one has $x_N = 0$ and, from $Ax = b$,

$$A_B x_B = b,$$

so $x_B = A_B^{-1}b \geq 0$. Thus x is a BFS. □

Theorem 1.3 (Vertex theorem). *If an optimal solution exists, then there exists an optimal solution that is a vertex of P .*

Proof. Let $\bar{x}$ be an optimal solution maximizing the number of active nonnegativity constraints. Suppose, by contradiction, that $\bar{x}$ is not a vertex. Then, by the vertex–BFS equivalence, $\bar{x}$ is not a basic feasible solution. Hence there exists a nonzero direction $d \in \mathbb{R}^n$ such that

$$Ad = 0, \quad d_j = 0 \quad \text{for every } j \in I(\bar{x}).$$

Moreover, by changing the sign of d if necessary, we may assume that

$$c^\top d \leq 0.$$

Consider

$$x(\lambda) = \bar{x} + \lambda d, \quad \lambda \geq 0.$$

For all λ , the equality constraint is preserved, since $Ad = 0$, and the constraints active at $\bar{x}$ remain active, since $d_j = 0$ for every $j \in I(\bar{x})$.

If $x(\lambda)$ is feasible for every $\lambda \geq 0$, then either $c^\top d < 0$, in which case the objective value tends to $-\infty$, contradicting optimality, or $c^\top d = 0$. In the latter case, choose the sign of d so that at least one positive component of $\bar{x}$ decreases along the direction d . Then there exists a finite $\lambda^* > 0$ such that $x(\lambda^*) \in P$ and at least one additional nonnegativity constraint becomes active. Since $c^\top d = 0$, this new point is still optimal, contradicting the maximality of the number of active constraints at $\bar{x}$.

Therefore $\bar{x}$ must be a vertex. Hence, if an optimal solution exists, there exists an optimal solution that is a vertex of P . □

Remark 1.4. By the previous theorem and the vertex–BFS equivalence, if an optimum exists, then there also exists an optimal BFS.

Theorem 1.5 (Fundamental theorem of the simplex method). *Let x^B be a basic feasible solution associated with a basis B , and let $k \in N$ be such that $\bar{c}_k < 0$. Define the direction $d^{(k)}$ by*

$$d_k^{(k)} = 1, \quad d_j^{(k)} = 0 \quad (j \in N \setminus \{k\}), \quad d_B^{(k)} = -A_B^{-1}A_k.$$

Then $Ad^{(k)} = 0$ and, for every feasible step $t > 0$,

$$c^\top(x^B + td^{(k)}) = c^\top x^B + t\bar{c}_k < c^\top x^B.$$

Hence either there exists a finite maximum step leading to a new better basic feasible solution, or the problem is unbounded below.

Proof. The direction $d^{(k)}$ is built by letting the nonbasic variable x_k enter the basis, keeping the other nonbasic variables fixed, and adjusting the basic variables so that the equality constraint $Ax = b$ remains satisfied. Indeed,

$$Ad^{(k)} = A_B(-A_B^{-1}A_k) + A_k = 0.$$

Thus

$$x(t) := x^B + td^{(k)}$$

satisfies $Ax(t) = b$ for all $t \geq 0$. Moreover, the basic components of $x(t)$ are

$$x_B(t) = x_B^B - tA_B^{-1}A_k,$$

while the nonbasic components are fixed, except for $x_k(t) = t$. Therefore

$$c^\top x(t) = c^\top x^B + t(c_k - c_B^\top A_B^{-1}A_k) = c^\top x^B + t\bar{c}_k.$$

Since $\bar{c}_k < 0$, every feasible step $t > 0$ strictly improves the objective value.

If there exists $i \in B$ such that $(A_B^{-1}A_k)_i > 0$, then feasibility requires

$$t \leq \frac{x_i^B}{(A_B^{-1}A_k)_i}.$$

Hence the maximum feasible step is

$$t^* = \min_{i \in B: (A_B^{-1}A_k)_i > 0} \frac{x_i^B}{(A_B^{-1}A_k)_i}.$$

At $t = t^*$, at least one basic variable becomes zero. Thus one basic variable leaves the basis, x_k enters the basis, and a new basic feasible solution is obtained. Since $t^* > 0$ and $\bar{c}_k < 0$, this new basic feasible solution has strictly smaller objective value.

If instead $A_B^{-1}A_k \leq 0$ componentwise, then

$$x_B(t) = x_B^B - tA_B^{-1}A_k \geq x_B^B \geq 0$$

for every $t \geq 0$. Hence $x(t)$ is feasible for every $t \geq 0$. Since

$$c^\top x(t) = c^\top x^B + t\bar{c}_k$$

and $\bar{c}_k < 0$, we have $c^\top x(t) \rightarrow -\infty$ as $t \rightarrow +\infty$. Therefore the problem is unbounded below. $\square$

Theorem 1.6 (Optimality condition). *Let x^B be a BFS associated with basis B . If $\bar{c}_j \geq 0$ for every $j \in N$, then x^B is optimal.*

Proof. For every $x \in P$, write $x = (x_B, x_N)$. From

$$A_B x_B + A_N x_N = b$$

one gets

$$x_B = A_B^{-1}b - A_B^{-1}A_N x_N = x_B^B - A_B^{-1}A_N x_N.$$

Substituting into the objective function,

$$\begin{aligned} c^\top x &= c_B^\top x_B + c_N^\top x_N \\ &= c_B^\top (x_B^B - A_B^{-1}A_N x_N) + c_N^\top x_N \\ &= c^\top x^B + \sum_{j \in N} (c_j - c_B^\top A_B^{-1}A_j) x_j \\ &= c^\top x^B + \sum_{j \in N} \bar{c}_j x_j. \end{aligned}$$

If $x_j \geq 0$ and $\bar{c}_j \geq 0$ for every $j \in N$, then $c^\top x \geq c^\top x^B$. Hence x^B is optimal. $\square$

Theorem 1.7 (Unboundedness theorem). *If there exists $k \in N$ such that $\bar{c}_k < 0$ and $A_B^{-1}A_k \leq 0$, then the problem is unbounded below.*

Proof. This is exactly the second case of the fundamental theorem of the simplex method: along the direction $d^{(k)}$, all points $x^B + td^{(k)}$ remain feasible and the objective value tends to $-\infty$. $\square$

Theorem 1.8 (Existence theorem). *For a linear program, exactly one of the following three cases occurs:*

1. an optimal solution exists;
2. the problem is unbounded below;
3. the problem is infeasible.

Proof. If $P = \emptyset$, the problem is infeasible. Assume now that $P \neq \emptyset$.

If the objective function is bounded below on P , then, since P is a polyhedron and the objective function is linear, the infimum is attained on a face of P . In particular, the problem admits an optimal solution.

If, on the other hand, the objective function is not bounded below on P , then for every $M \in \mathbb{R}$ there exists $x \in P$ such that

$$c^\top x < M.$$

Hence the problem is unbounded below.

Therefore exactly one of the following alternatives holds: the problem is infeasible, the problem has an optimal solution, or the problem is unbounded below. These alternatives are mutually exclusive. $\square$

1.4 The simplex algorithm

Starting from a feasible basis B , one simplex iteration is:

1. compute $x_B = A_B^{-1}b$ and set $x_N = 0$;
2. compute the reduced costs $\bar{c}_j = c_j - c_B^\top A_B^{-1}A_j$ for $j \in N$;
3. if $\bar{c}_j \geq 0$ for all $j \in N$, stop: the current BFS is optimal;
4. otherwise choose $j^* \in N$ with $\bar{c}_{j^*} < 0$;

5. compute $u = A_B^{-1} A_{j^*}$;
6. if $u_i \leq 0$ for every i , stop: the problem is unbounded below;
7. otherwise perform the ratio test

$$\theta^* = \min_{i: u_i > 0} \frac{(x_B)_i}{u_i};$$

8. the index attaining the minimum leaves the basis, while j^* enters the basis.

The process is repeated until optimality or unboundedness is detected.

2 Duality and complementary slackness

In the previous section we developed the simplex method as a constructive algorithm: starting from a basic feasible solution, we moved to neighboring vertices of the feasible polyhedron until the optimality condition $\bar{c}_j \geq 0$ for all nonbasic indices was satisfied. That condition was justified by an algebraic substitution argument, but it leaves a more fundamental question open: why is nonnegativity of the reduced costs enough to certify optimality?

The answer is duality. To every linear program one can associate a second linear program, its dual, whose feasible solutions act as certificates for the first. Weak duality says that any feasible dual solution gives a one-sided bound on the primal value; strong duality says that, under the usual feasibility and boundedness assumptions, the best such bound is exact; complementary slackness says that optimal primal and dual solutions are linked by algebraic equations that identify where the primal variables can be positive.

This section develops these ideas in the same standard-form minimization setting used above. The reason for doing so is not only conceptual. In the next sections, flow problems and cuts will appear as a primal-dual pair, while the discrete Kantorovich problem will give dual variables with a concrete interpretation as transport potentials. Complementary slackness will then become the statement that mass is transported only along tight edges.

2.1 The dual linear program

We work with the standard-form primal problem

$$(P) \quad p^* := \min\{c^\top x : Ax = b, x \geq 0\},$$

where $A \in \mathbb{R}^{m \times n}$ has rank m . The convention is that $p^* = +\infty$ if (P) is infeasible and $p^* = -\infty$ if it is unbounded below.

Duality can be motivated by asking how to certify a lower bound on p^* . Fix any vector $y \in \mathbb{R}^m$. For every primal feasible x ,

$$y^\top Ax = y^\top b = b^\top y.$$

If, in addition,

$$A^\top y \leq c \quad \text{componentwise,}$$

then for every $x \geq 0$,

$$b^\top y = y^\top Ax = (A^\top y)^\top x \leq c^\top x.$$

Thus $b^\top y$ is a lower bound on every feasible primal value. The best lower bound obtained in this way is the dual problem

$$(D) \quad d^* := \max\{b^\top y : A^\top y \leq c\}, \quad y \in \mathbb{R}^m.$$

The variables y are free in sign because the primal constraints are equalities. They are often called dual variables, prices, or potentials. There are m dual variables, one per primal equality constraint, and n dual inequalities, one per primal variable.

This viewpoint also explains the dual interpretation of reduced costs. Given a basis B , define

$$y^\top := c_B^\top A_B^{-1}.$$

Then

$$\bar{c}_j = c_j - c_B^\top A_B^{-1} A_j = c_j - A_j^\top y.$$

Therefore the simplex optimality condition $\bar{c}_j \geq 0$ for every j is exactly the dual feasibility condition $A^\top y \leq c$.

2.2 Weak duality

Theorem 2.1 (Weak duality). *If x is feasible for (P) and y is feasible for (D), then*

$$b^\top y \leq c^\top x.$$

Consequently, $d^* \leq p^*$.

Proof. Since $Ax = b$,

$$b^\top y = y^\top b = y^\top Ax = (A^\top y)^\top x.$$

Since $A^\top y \leq c$ and $x \geq 0$,

$$(A^\top y)^\top x \leq c^\top x.$$

Thus $b^\top y \leq c^\top x$. Taking the supremum over all dual feasible y and the infimum over all primal feasible x gives $d^* \leq p^*$. $\square$

Corollary 2.2. *For the pair (P)–(D):*

1. *if (P) is unbounded below, then (D) is infeasible;*
2. *if (D) is unbounded above, then (P) is infeasible;*
3. *if x is feasible for (P), y is feasible for (D), and $c^\top x = b^\top y$, then x is optimal for (P) and y is optimal for (D).*

Proof. The first two statements follow directly from weak duality, because a feasible point on the opposite side would give a finite one-sided bound. For the third statement, weak duality gives

$$b^\top y \leq p^* \leq c^\top x.$$

If $b^\top y = c^\top x$, then all inequalities are equalities, so both feasible points are optimal. $\square$

Weak duality is already useful: it turns any dual feasible vector into a certificate that no primal feasible solution can be cheaper than $b^\top y$. Strong duality says that, when the primal has a finite optimum, there is a certificate whose value is exactly the optimum.

2.3 Geometric prerequisites: projection and separation

A proof of strong duality can be obtained from the simplex method, but it is useful first to isolate the geometric principle behind it. The key idea is separation: a point outside a closed convex set can be strictly separated from that set by a hyperplane.

Theorem 2.3 (Projection onto a closed convex set). *Let $\mathcal{C} \subseteq \mathbb{R}^k$ be closed and convex, and let $\xi \in \mathbb{R}^k \setminus \mathcal{C}$. There exists a unique point $x^* \in \mathcal{C}$ such that*

$$\|x^* - \xi\| \leq \|x - \xi\| \quad \text{for every } x \in \mathcal{C}.$$

Proof. Let $\delta = \inf_{x \in \mathcal{C}} \|x - \xi\|^2$ and let (x_n) be a minimizing sequence. By the parallelogram identity, for any m, n ,

$$\|x_m - x_n\|^2 = 2\|x_m - \xi\|^2 + 2\|x_n - \xi\|^2 - 4\left\|\frac{x_m + x_n}{2} - \xi\right\|^2.$$

Since $\mathcal{C}$ is convex, $(x_m + x_n)/2 \in \mathcal{C}$, and therefore the last squared norm is at least δ . Hence (x_n) is Cauchy. It converges to some x^* , and closedness gives $x^* \in \mathcal{C}$. By continuity, x^* attains the minimum. Uniqueness follows from the same parallelogram identity applied to two distinct minimizers. $\square$

Theorem 2.4 (Separating hyperplane). *Let $\mathcal{C} \subseteq \mathbb{R}^k$ be closed and convex, and let $\xi \in \mathbb{R}^k \setminus \mathcal{C}$. There exist a vector $\lambda \in \mathbb{R}^k$ and a constant $\delta > 0$ such that*

$$\lambda^\top(x - \xi) \geq \delta \quad \text{for every } x \in \mathcal{C}.$$

Proof. Let x^* be the projection of ξ onto $\mathcal{C}$ and set $\lambda = x^* - \xi$. Then $\lambda \neq 0$. For any $x \in \mathcal{C}$ and $\theta \in (0, 1)$, convexity gives $(1 - \theta)x^* + \theta x \in \mathcal{C}$. Hence

$$\|(1 - \theta)x^* + \theta x - \xi\|^2 \geq \|x^* - \xi\|^2.$$

Expanding and letting $\theta \downarrow 0$ gives $\lambda^\top(x - x^*) \geq 0$. Therefore

$$\lambda^\top(x - \xi) = \lambda^\top(x - x^*) + \lambda^\top(x^* - \xi) \geq \|\lambda\|^2.$$

Taking $\delta = \|\lambda\|^2 > 0$ proves the claim. $\square$

We will also use the following standard fact.

Theorem 2.5 (Linear image of the nonnegative orthant). *Let M be a real $k \times h$ matrix. The set*

$$\mathcal{C} = \{Mz : z \in \mathbb{R}^h, z \geq 0\}$$

is closed and convex.

Idea of proof. Convexity is immediate. Closedness follows from the finite-dimensional conic structure of the nonnegative orthant: by Caratheodory's theorem, every point of $\mathcal{C}$ can be represented using at most k linearly independent columns of M . There are only finitely many such subsets of columns, and on each fixed linearly independent subset the corresponding cone is closed. The finite union of these closed cones is closed. $\square$

2.4 Strong duality

Theorem 2.6 (Strong duality). *Assume that the primal problem (P) is feasible and has a finite optimal value. Then the dual problem (D) has an optimal solution y^* and*

$$\min(P) = \max(D).$$

Proof. We give a proof based on the simplex optimality condition, which is the most natural in these notes. Since the primal problem is feasible and has a finite optimal value, the vertex theorem gives an optimal basic feasible solution. Choose an optimal basis B satisfying the simplex optimality condition

$$\bar{c}_j = c_j - c_B^\top A_B^{-1} A_j \geq 0 \quad \text{for every } j \in N.$$

Define

$$(y^*)^\top := c_B^\top A_B^{-1}.$$

Then, for every column A_j of A ,

$$c_j - A_j^\top y^* = c_j - c_B^\top A_B^{-1} A_j \geq 0.$$

Hence $A^\top y^* \leq c$, so y^* is feasible for the dual problem.

Moreover, since $x_N^* = 0$ and $x_B^* = A_B^{-1}b$,

$$c^\top x^* = c_B^\top x_B^* = c_B^\top A_B^{-1}b = (y^*)^\top b = b^\top y^*.$$

We have therefore exhibited a primal feasible point and a dual feasible point with the same value. By weak duality, both are optimal and the two optimal values coincide. $\square$

The proof shows that the dual optimum is not an abstract object: at termination of the simplex method, it is obtained directly from the optimal basis through

$$y^\top = c_B^\top A_B^{-1}.$$

This is why the reduced costs are dual slacks. Indeed, $\bar{c}_j = c_j - A_j^\top y$, so a zero reduced cost is a tight dual constraint.

2.5 Complementary slackness

Strong duality identifies the optimal values. Complementary slackness identifies the relation between optimal solutions themselves.

Theorem 2.7 (Complementary slackness). *Let x be feasible for (P) and y feasible for (D) . The following are equivalent:*

(i) x is optimal for (P) and y is optimal for (D) ;

(ii) $x^\top(c - A^\top y) = 0$;

(iii) for every $j = 1, \dots, n$,

$$x_j(c_j - A_j^\top y) = 0,$$

equivalently, either $x_j = 0$ or $A_j^\top y = c_j$.

Proof. The duality gap is

$$c^\top x - b^\top y = c^\top x - y^\top A x = x^\top(c - A^\top y) = \sum_{j=1}^n x_j(c_j - A_j^\top y).$$

Each term is nonnegative because $x_j \geq 0$ and $c_j - A_j^\top y \geq 0$ by dual feasibility. Hence the gap is zero if and only if every term is zero. By weak duality, zero gap is equivalent to simultaneous optimality. $\square$

The componentwise form is the one that will be used later. It says that a primal variable can be positive only when the corresponding dual inequality is tight. In transport language, a shipment p_{ij} can be positive only if the corresponding edge (i, j) is tight, meaning that the transport cost is exactly explained by the two dual potentials.

2.6 Duality for general linear programs

The standard-form pair used so far is sufficient for the theoretical development, but it is useful to record the general correspondence. For a primal minimization problem, the dual is a maximization problem, and constraints and variables exchange roles according to the following table:

Primal minimization element	Dual maximization counterpart
$\sum_j a_{ij}x_j \geq b_i$	$y_i \geq 0$
$\sum_j a_{ij}x_j \leq b_i$	$y_i \leq 0$
$\sum_j a_{ij}x_j = b_i$	y_i free
$x_j \geq 0$	$\sum_i a_{ij}y_i \leq c_j$
$x_j \leq 0$	$\sum_i a_{ij}y_i \geq c_j$
x_j free	$\sum_i a_{ij}y_i = c_j$

For a maximization problem, the same table is read in the opposite direction. The standard-form minimization pair is therefore

$$\min\{c^T x : Ax = b, x \geq 0\} \iff \max\{b^T y : A^T y \leq c, y \in \mathbb{R}^m\}.$$

This is exactly the form of the discrete Kantorovich problem: the equality constraints are marginal constraints, and the free dual variables become the Kantorovich potentials.

2.7 A worked example

Consider the canonical-form primal problem

$$(P) \quad \max 5x_1 + 4x_2$$

subject to

$$x_1 \leq 4,$$

$$x_1 + 2x_2 \leq 10,$$

$$3x_1 + 2x_2 \leq 16,$$

$$x_1, x_2 \geq 0.$$

Its dual is

$$(D) \quad \min 4y_1 + 10y_2 + 16y_3$$

subject to

$$y_1 + y_2 + 3y_3 \geq 5,$$

$$2y_2 + 2y_3 \geq 4,$$

$$y_1, y_2, y_3 \geq 0.$$

Take the primal candidate

$$x = (3, 3.5).$$

It is feasible because

$$x_1 = 3 < 4, \quad x_1 + 2x_2 = 10, \quad 3x_1 + 2x_2 = 16.$$

Its value is

$$5 \cdot 3 + 4 \cdot 3.5 = 29.$$

Complementary slackness can be used to recover a dual certificate. Since

$$x_1 > 0, \quad x_2 > 0,$$

both dual inequalities must be tight:

$$y_1 + y_2 + 3y_3 = 5,$$

$$2y_2 + 2y_3 = 4.$$

Since the first primal constraint is slack, the corresponding dual variable must be zero:

$$y_1 = 0.$$

Solving gives

$$y_1 = 0, \quad y_2 = \frac{1}{2}, \quad y_3 = \frac{3}{2}.$$

This vector is dual feasible and has value

$$4 \cdot 0 + 10 \cdot \frac{1}{2} + 16 \cdot \frac{3}{2} = 29.$$

The primal and dual feasible values coincide. Hence both candidates are optimal.

2.8 Outlook: from duality to flows and discrete OT

The rest of the notes uses duality in two structured settings. In network flows, dual variables attached to vertex constraints lead to cuts, and weak duality becomes the statement that every flow is bounded above by every cut capacity. The max-flow/min-cut theorem is the corresponding strong-duality statement.

In discrete optimal transport, the primal variables are the masses p_{ij} transported from source i to target j . The equality constraints impose the source and target marginals, so the dual variables are two families of free potentials, u_i and v_j . The dual constraint is

$$u_i + v_j \leq c_{ij},$$

and complementary slackness becomes

$$p_{ij} > 0 \implies u_i + v_j = c_{ij}.$$

This is the same tight-edge principle used by primal-dual algorithms such as the Hungarian method and by the transportation simplex.

3 Network flows as linear programs

Before moving to optimal transport, it is useful to look at flow problems. They are structured LPs in which variables represent quantities moving along arcs of a directed graph.

Let $D = (V, A)$ be a directed graph with source s , sink t , and capacity function $u : A \rightarrow \mathbb{R}_+$. For each arc $a \in A$, let f_a denote the amount of flow sent through a . A feasible s - t flow satisfies

$$0 \leq f_a \leq u_a \quad (a \in A)$$

and the conservation constraints

$$\sum_{a \in \delta^{\text{in}}(v)} f_a = \sum_{a \in \delta^{\text{out}}(v)} f_a \quad (v \in V \setminus \{s, t\}).$$

The value of the flow is

$$\text{val}(f) = \sum_{a \in \delta^{\text{out}}(s)} f_a - \sum_{a \in \delta^{\text{in}}(s)} f_a.$$

Thus the maximum-flow problem is the LP

$$\max \text{val}(f) \quad \text{subject to conservation and capacity constraints.}$$

The dual viewpoint appears through cuts. If $U \subseteq V$ with $s \in U$ and $t \notin U$, the capacity of the cut is

$$u(\delta^{\text{out}}(U)) = \sum_{a \in \delta^{\text{out}}(U)} u_a.$$

Every feasible flow satisfies

$$\text{val}(f) \leq u(\delta^{\text{out}}(U)),$$

so cuts are dual upper bounds on flows. The max-flow/min-cut theorem states that the maximum value of an s - t flow equals the minimum capacity of an s - t cut. Algorithms such as Ford-Fulkerson make this equality constructive by repeatedly augmenting flow along paths in the residual graph.

This is the bridge to optimal transport: both flow problems and OT describe movement of quantities by linear constraints. However, maximum flow asks how much can be sent through a network, while OT fixes the total mass and minimizes the cost of moving it from one distribution to another.

4 The discrete Kantorovich problem as an LP

Let

$$a = (a_1, \dots, a_m), \quad b = (b_1, \dots, b_n)$$

be two discrete probability distributions, so $a_i, b_j \geq 0$ and

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j = 1.$$

Let $C = (c_{ij})$ be a cost matrix, where c_{ij} is the cost of transporting one unit of mass from source point i to target point j .

A transport plan is a matrix $P = (p_{ij})$ with $p_{ij} \geq 0$, where p_{ij} represents the amount of mass moved from i to j . The marginal constraints are

$$\sum_{j=1}^n p_{ij} = a_i \quad (i = 1, \dots, m),$$

$$\sum_{i=1}^m p_{ij} = b_j \quad (j = 1, \dots, n).$$

The discrete Kantorovich problem is therefore the linear program

$$\begin{aligned} \min_{P \geq 0} \quad & \sum_{i=1}^m \sum_{j=1}^n c_{ij} p_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^n p_{ij} = a_i \quad (i = 1, \dots, m), \\ & \sum_{i=1}^m p_{ij} = b_j \quad (j = 1, \dots, n). \end{aligned}$$

4.1 Dual formulation

The dual variables are potentials u_i for the source constraints and v_j for the target constraints. Since the marginal constraints are equalities, these dual variables are free in sign. The dual problem is

$$\begin{aligned} \max_{u,v} \quad & \sum_{i=1}^m a_i u_i + \sum_{j=1}^n b_j v_j \\ \text{s.t.} \quad & u_i + v_j \leq c_{ij} \quad \text{for all } i, j. \end{aligned}$$

Weak duality says that any potentials satisfying $u_i + v_j \leq c_{ij}$ give a lower bound on the cost of any feasible transport plan. Strong duality says that, under feasibility, the minimum transport cost equals the maximum dual value.

Complementary slackness becomes especially intuitive:

$$p_{ij} > 0 \implies u_i + v_j = c_{ij}.$$

Mass is transported only along tight edges, where the transport cost is exactly explained by the two potentials. This is the same principle used by primal-dual algorithms: maintain dual feasibility, identify tight edges, and construct a primal solution supported on them.

5 Birkhoff-von Neumann theorem

Birkhoff-von Neumann's theorem is naturally related to discrete OT. If the two marginals are uniform on n points, then a transport plan can be represented, after multiplying by n , as a doubly stochastic matrix. Permutation matrices correspond to deterministic one-to-one transport plans. The theorem says that every such randomized plan is a convex combination of deterministic matchings.

Definition 5.1. A square matrix is doubly stochastic if all entries are nonnegative and every row and every column sums to one.

Definition 5.2. A permutation matrix is a 0-1 matrix with exactly one entry equal to 1 in each row and in each column.

Theorem 5.3 (Birkhoff-von Neumann). *Every doubly stochastic matrix is a convex combination of permutation matrices.*

Proof. We use Hall's theorem. Given a doubly stochastic matrix $M = (m_{ij})$, construct a bipartite graph with one vertex for each row and one vertex for each column, and draw an edge $i \sim j$ whenever $m_{ij} > 0$.

First we show that this graph has a perfect matching. Let S be a set of row vertices and let $R(S)$ be its set of neighboring column vertices. Since all nonzero entries in the rows indexed by S lie in the columns $R(S)$,

$$\sum_{i \in S} \sum_{j \in R(S)} m_{ij} = \sum_{i \in S} \sum_{j=1}^n m_{ij} = |S|.$$

On the other hand, because each column sum is one,

$$\sum_{i \in S} \sum_{j \in R(S)} m_{ij} \leq \sum_{j \in R(S)} \sum_{i=1}^n m_{ij} = |R(S)|.$$

Hence $|S| \leq |R(S)|$, so Hall's condition holds. Therefore the graph has a perfect matching.

Let P_0 be the permutation matrix associated with such a matching and define

$$\delta_0 := \min\{m_{ij} : (P_0)_{ij} = 1\} > 0.$$

If $\delta_0 = 1$, then $M = P_0$. Otherwise set

$$M_1 := \frac{M - \delta_0 P_0}{1 - \delta_0}.$$

The matrix M_1 is again doubly stochastic: nonnegativity follows from the definition of δ_0 , and each row and column sum is divided by $1 - \delta_0$ after being reduced by exactly δ_0 . Moreover, at least one positive entry of M becomes zero, so M_1 has fewer nonzero entries than M .

Repeating the argument, or equivalently applying induction on the number of nonzero entries, we eventually obtain permutation matrices $P_0, P_1, \dots, P_r$ and nonnegative coefficients whose sum is one such that

$$M = \lambda_0 P_0 + \lambda_1 P_1 + \dots + \lambda_r P_r.$$

Thus M is a convex combination of permutation matrices. □

For completeness, we recall the matching result used above.

Theorem 5.4 (Hall's theorem). *Let $G = (A \cup B, E)$ be a bipartite graph. There is a matching saturating A if and only if every subset $S \subseteq A$ has at least $|S|$ neighbors in B .*

Proof. If a matching saturating A exists, then the vertices matched to the elements of S are distinct neighbors of S , so Hall's condition is necessary.

Conversely, assume Hall's condition and prove the result by induction on $|A|$. The case $|A| = 1$ is immediate. If every proper nonempty subset $S \subset A$ has strictly more than $|S|$ neighbors, choose any edge xy with $x \in A$ and $y \in B$. After removing x and y , Hall's condition still holds, so induction gives a matching of $A \setminus \{x\}$; adding xy completes the matching.

Otherwise, there exists a proper nonempty set $S \subset A$ whose neighbor set T satisfies $|T| = |S|$. By induction, the subgraph induced by $S \cup T$ has a matching saturating S . For any $S' \subseteq A \setminus S$, Hall's condition applied to $S \cup S'$ implies that S' has at least $|S'|$ neighbors outside T . Hence Hall's condition also holds for the subgraph induced by $(A \setminus S) \cup (B \setminus T)$, so induction gives a matching saturating $A \setminus S$. Combining the two matchings gives a matching saturating all of A . □

6 Algorithms for discrete optimal transport

6.1 Hungarian algorithm

The Hungarian algorithm solves the square assignment problem

$$\begin{aligned} \min \quad & \sum_{i,j} c_{ij} x_{ij} \\ \text{s.t.} \quad & \sum_j x_{ij} = 1 \quad \text{for every } i, \\ & \sum_i x_{ij} = 1 \quad \text{for every } j, \\ & x_{ij} \in \{0, 1\}. \end{aligned}$$

This is the problem of finding a minimum-cost perfect matching in a bipartite graph. It is also a special case of discrete OT with uniform marginals and deterministic transport plans.

The primal-dual structure is the key point. Introduce dual variables u_i for rows and v_j for columns. The dual constraints are

$$u_i + v_j \leq c_{ij}.$$

The reduced costs are

$$\bar{c}_{ij} = c_{ij} - u_i - v_j \geq 0.$$

The equality graph is the graph of tight edges, namely those with $\bar{c}_{ij} = 0$. If this graph contains a perfect matching, then the matching is optimal by complementary slackness.

A standard initialization is

$$u_i = \min_j c_{ij}, \quad v_j = 0,$$

which guarantees dual feasibility and at least one tight edge in each row. If the current equality graph does not contain a perfect matching, the algorithm builds an alternating tree from unmatched rows. Let S be the reachable rows and T the reachable columns. Define

$$\delta = \min\{\bar{c}_{ij} : i \in S, j \notin T\}.$$

Then update

$$u_i \leftarrow u_i + \delta \quad (i \in S), \quad v_j \leftarrow v_j - \delta \quad (j \in T).$$

This preserves dual feasibility and creates at least one new tight edge, allowing the matching to grow. The algorithm terminates when a perfect matching is found.

6.2 Dantzig's transportation simplex

The transportation problem is the LP

$$\begin{aligned} \min \quad & \sum_{i,j} c_{ij} p_{ij}, \\ \text{s.t.} \quad & \sum_j p_{ij} = a_i, \\ & \sum_i p_{ij} = b_j, \\ & p_{ij} \geq 0. \end{aligned}$$

Dantzig's transportation simplex is the simplex method specialized to this network structure. A basic feasible transport plan corresponds to a spanning tree in the bipartite graph of sources and targets, typically with $m + n - 1$ basic variables in the nondegenerate case.

The dual potentials u_i, v_j satisfy

$$u_i + v_j = c_{ij} \quad \text{on basic arcs.}$$

For nonbasic arcs, the reduced cost is

$$\bar{c}_{ij} = c_{ij} - u_i - v_j.$$

If all reduced costs are nonnegative, the current plan is optimal. If some $\bar{c}_{ij} < 0$, that arc enters the basis. Adding it creates a unique cycle in the bipartite graph; moving mass alternately along the cycle preserves the marginals. The maximum feasible move is determined by the smallest mass on the subtracting positions of the cycle, and one arc leaves the basis. This is exactly the simplex pivot, but expressed in transport language.

7 Conclusion

These notes have presented linear programming as a unifying language for several problems in discrete optimization and optimal transport. Starting from the geometry of polyhedra and the simplex method, we have seen how basic feasible solutions, reduced costs, and optimality conditions naturally lead to duality and complementary slackness. This primal–dual viewpoint is essential because it turns optimality into a certificate: feasible dual variables provide bounds, and equality of primal and dual values proves optimality.

The same ideas appear in network flows, assignment problems, and discrete optimal transport. In the Kantorovich formulation, transport plans are feasible points of a linear program, while the dual variables become potentials satisfying

$$u_i + v_j \leq c_{ij}.$$

Complementary slackness then gives the fundamental tight-edge condition: mass can be transported only along pairs for which the dual inequality is saturated. This principle also explains the structure of algorithms such as the Hungarian method and the transportation simplex.

Finally, the Birkhoff–von Neumann theorem shows that, in the case of uniform marginals, doubly stochastic transport plans can be decomposed into convex combinations of permutation matrices. Thus deterministic matchings are the extreme building blocks of randomized transport plans. Overall, linear programming provides both the theoretical foundation and the algorithmic framework for understanding discrete optimal transport and its main computational methods.

References

- [1] R. J. Vanderbei, *Linear Programming: Foundations and Extensions*. Springer, 2020.
- [2] G. Peyré and M. Cuturi, *Computational Optimal Transport*. Foundations and Trends in Machine Learning, 2019.
- [3] H. W. Kuhn, The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 1955.
- [4] M. Tehranchi, *Notes on the Birkhoff–von Neumann theorem*.
- [5] G. Hurlbert, A short proof of the Birkhoff–von Neumann theorem.