Part IB — Optimisation

Theorems with proof

Based on lectures by F. A. Fischer

Notes taken by Dexter Chua

Easter 2015

Lagrangian methods

General formulation of constrained problems; the Lagrangian sufficiency theorem. Interpretation of Lagrange multipliers as shadow prices. Examples. [2]

Linear programming in the nondegenerate case

Convexity of feasible region; sufficiency of extreme points. Standardization of problems, slack variables, equivalence of extreme points and basic solutions. The primal simplex algorithm, artificial variables, the two-phase method. Practical use of the algorithm; the tableau. Examples. The dual linear problem, duality theorem in a standardized case, complementary slackness, dual variables and their interpretation as shadow prices. Relationship of the primal simplex algorithm to dual problem. Two person zero-sum games. [6]

Network problems

The Ford-Fulkerson algorithm and the max-flow min-cut theorems in the rational case. Network flows with costs, the transportation algorithm, relationship of dual variables with nodes. Examples. Conditions for optimality in more general networks; *the simplex-on-a-graph algorithm*. [3]

Practice and applications

Efficiency of algorithms. The formulation of simple practical and combinatorial problems as linear programming or network problems. [1]

Contents

1 Introduction and preliminaries

1.1 Constrained optimization

1.2 Review of unconstrained optimization

Lemma. Let f be twice differentiable. Then f is convex on a convex set S if the Hessian matrix

$$Hf_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

is positive semidefinite for all $x \in S$, where this fancy term means:

Theorem. Let $X \subseteq \mathbb{R}^n$ be convex, $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable on X. If $x^* \in X$ satisfy $\nabla f(x^*) = 0$ and $Hf(x)$ is positive semidefinite for all $x \in X$, then x^* minimizes f on X.

2 The method of Lagrange multipliers

Theorem (Lagrangian sufficiency)**.** Let $x^* \in X$ and $\lambda^* \in \mathbb{R}^m$ be such that

$$L(x^*, \lambda^*) = \inf_{x \in X} L(x, \lambda^*) \quad \text{and} \quad h(x^*) = b.$$

Then x^* is optimal for (P).

In words, if x^* minimizes L for a fixed λ^*, and x^* satisfies the constraints, then x^* minimizes f.

Proof. We first define the "feasible set": let $X(b) = \{x \in X : h(x) = b\}$, i.e. the set of all x that satisfies the constraints. Then

$$\begin{aligned}
\min_{x \in X(b)} f(x) &= \min_{x \in X(b)} (f(x) - \lambda^{*T}(h(x) - b)) \quad \text{since } h(x) - b = 0\\
&\geq \min_{x \in X} (f(x) - \lambda^{*T}(h(x) - b))\\
&= f(x^*) - \lambda^{*T}(h(x^*) - b).\\
&= f(x^*).
\end{aligned}$$

□

2.1 Complementary Slackness

2.2 Shadow prices

Theorem. Consider the problem

$$\text{minimize } f(x) \text{ subject to } h(x) = b.$$

Here we assume all functions are continuously differentiable. Suppose that for each $b \in \mathbb{R}^n$, $\phi(b)$ is the optimal value of f and λ^* is the corresponding Lagrange multiplier. Then

$$\frac{\partial \phi}{\partial b_i} = \lambda_i^*.$$

2.3 Lagrange duality

Theorem (Weak duality)**.** If $x \in X(b)$ (i.e. x satisfies both the functional and regional constraints) and $\lambda \in Y$, then

$$g(\lambda) \leq f(x).$$

In particular,

$$\sup_{\lambda \in Y} g(\lambda) \leq \inf_{x \in X(b)} f(x).$$

Proof.

$$\begin{aligned}
g(\lambda) &= \inf_{x' \in X} L(x', \lambda)\\
&\leq L(x, \lambda)\\
&= f(x) - \lambda^T(h(x) - b)\\
&= f(x).
\end{aligned}$$

□

2.4 Supporting hyperplanes and convexity

Theorem. (P) satisfies strong duality iff $\phi(c) = \inf_{x \in X(c)} f(x)$ has a supporting hyperplane at b.

Proof. ($\Leftarrow$) Suppose there is a supporting hyperplane. Then since the plane passes through $\phi(b)$, it must be of the form

$$\alpha(c) = \phi(b) + \lambda^T(c - b).$$

Since this is supporting, for all $c \in \mathbb{R}^m$,

$$\phi(b) + \lambda^T(c - b) \leq \phi(c),$$

or

$$\phi(b) \leq \phi(c) - \lambda^T(c - b),$$

This implies that

$$\begin{aligned}\phi(b) &\leq \inf_{c \in \mathbb{R}^m} (\phi(c) - \lambda^T(c - b)) \\ &= \inf_{c \in \mathbb{R}^m} \inf_{x \in X(c)} (f(x) - \lambda^T(h(x) - b))\end{aligned}$$

(since $\phi(c) = \inf_{x \in X(c)} f(x)$ and $h(x) = c$ for $x \in X(c)$)

$$= \inf_{x \in X} L(x, \lambda).$$

(since $\bigcup_{c \in \mathbb{R}^m} X(c) = X$, which is true since for any $x \in X$, we have $x \in X(h(x))$)

$$= g(\lambda)$$

By weak duality, $g(\lambda) \leq \phi(b)$. So $\phi(b) = g(\lambda)$. So strong duality holds.

($\Rightarrow$). Assume now that we have strong duality. The there exists λ such that for all $c \in \mathbb{R}^m$,

$$\begin{aligned}\phi(b) &= g(\lambda) \\ &= \inf_{x \in X} L(x, \lambda) \\ &\leq \inf_{x \in X(c)} L(x, \lambda) \\ &= \inf_{x \in X(c)} (f(x) - \lambda^T(h(x) - b)) \\ &= \phi(c) - \lambda^T(c - b)\end{aligned}$$

So $\phi(b) + \lambda^T(c - b) \leq \phi(c)$. So this defines a supporting hyperplane. □

Theorem (Supporting hyperplane theorem)**.** Suppose that $\phi : \mathbb{R}^m \to \mathbb{R}$ is convex and $b \in \mathbb{R}^m$ lies in the interior of the set of points where ϕ is finite. Then there exists a supporting hyperplane to ϕ at b.

Theorem. Let

$$\phi(b) = \inf_{x \in X} \{f(x) : h(x) \leq b\}$$

If X, f, h are convex, then so is ϕ (assuming feasibility and boundedness).

Proof. Consider $b_1, b_2 \in \mathbb{R}^m$ such that $\phi(b_1)$ and $\phi(b_2)$ are defined. Let $\delta \in [0, 1]$ and define $b = \delta b_1 + (1-\delta)b_2$. We want to show that $\phi(b) \leq \delta\phi(b_1) + (1-\delta)\phi(b_2)$.

Consider $x_1 \in X(b_1)$, $x_2 \in X(b_2)$, and let $x = \delta x_1 + (1-\delta)x_2$. By convexity of X, $x \in X$.

By convexity of h,

$$\begin{aligned} h(x) &= h(\delta x_1 + (1-\delta)x_2) \\ &\leq \delta h(x_1) + (1-\delta)h(x_2) \\ &\leq \delta b_1 + (1-\delta)b_2 \\ &= b \end{aligned}$$

So $x \in X(b)$. Since $\phi(x)$ is an optimal solution, by convexity of f,

$$\begin{aligned} \phi(b) &\leq f(x) \\ &= f(\delta x_1 + (1-\delta)x_2) \\ &\leq \delta f(x_1) + (1-\delta)f(x_2) \end{aligned}$$

This holds for any $x_1 \in X(b_1)$ and $x_2 \in X(b_2)$. So by taking infimum of the right hand side,

$$\phi(b) \leq \delta\phi(b_1) + (1-\delta)\phi(b_2).$$

So ϕ is convex. □

Theorem. If a linear program is feasible and bounded, then it satisfies strong duality.

3 Solutions of linear programs

3.1 Linear programs

3.2 Basic solutions

Theorem. A vector x is a basic feasible solution of $Ax = b$ if and only if it is an extreme point of the set $X(b) = \{x' : Ax' = b, x' \geq 0\}$.

3.3 Extreme points and optimal solutions

Theorem. If (P) is feasible and bounded, then there exists an optimal solution that is a basic feasible solution.

Proof. Let x be optimal of (P). If x has at most non-zero entries, it is a basic feasible solution, and we are done.

Now suppose x has $r > m$ non-zero entries. Since it is not an extreme point, we have $y \neq z \in X(b)$, $\delta \in (0, 1)$ such that

$$x = \delta y + (1 - \delta)z.$$

We will show there exists an optimal solution strictly fewer than r non-zero entries. Then the result follows by induction.

By optimality of x, we have $c^T x \geq c^T y$ and $c^T x \geq c^T z$.

Since $c^T x = \delta c^T y + (1 - \delta)c^T z$, we must have that $c^T x = c^T y = c^T z$, i.e. y and z are also optimal.

Since $y \geq 0$ and $z \geq 0$, $x = \delta y + (1 - \delta)z$ implies that $y_i = z_i = 0$ whenever $x_i = 0$.

So the non-zero entries of y and z is a subset of the non-zero entries of x. So y and z have at most r non-zero entries, which must occur in rows where x is also non-zero.

If y or z has strictly fewer than r non-zero entries, then we are done. Otherwise, for any $\hat{\delta}$ (not necessarily in $(0, 1)$), let

$$x_{\hat{\delta}} = \hat{\delta}y + (1 - \hat{\delta})z = z + \hat{\delta}(y - z).$$

Observe that $x_{\hat{\delta}}$ is optimal for every $\hat{\delta} \in \mathbb{R}$.

Moreover, $y - z \neq 0$, and all non-zero entries of $y - z$ occur in rows where x is non-zero as well. We can thus choose $\hat{\delta} \in \mathbb{R}$ such that $x_{\hat{\delta}} \geq 0$ and $x_{\hat{\delta}}$ has strictly fewer than r non-zero entries. □

3.4 Linear programming duality

Theorem. The dual of the dual of a linear program is the primal.

Proof. It suffices to show this for the linear program in general form. We have shown above that the dual problem is

$$\text{minimize } -b^T \lambda \text{ subject to } -A^T \lambda \geq -c,\ \lambda \geq 0.$$

This problem has the same form as the primal, with $-b$ taking the role of c, $-c$ taking the role of b, $-A^T$ taking the role of A. So doing it again, we get back to the original problem. □

Theorem. Let x and λ be feasible for the primal and the dual of the linear program in general form. Then x and λ and optimal if and only if they satisfy complementary slackness, i.e. if

$$(c^T - \lambda^T A)x = 0 \text{ and } \lambda^T(Ax - b) = 0.$$

Proof. If x and λ are optimal, then

$$c^T x = \lambda^T b$$

since every linear program satisfies strong duality. So

$$\begin{aligned} c^T x &= \lambda^T b \\ &= \inf_{x' \in X}(c^T x' - \lambda^T(Ax' - b)) \\ &\leq c^T x - \lambda^T(Ax - b) \\ &\leq c^T x. \end{aligned}$$

The last line is since $Ax \geq b$ and $\lambda \geq 0$.

The first and last term are the same. So the inequalities hold with equality. Therefore

$$\lambda^T b = c^T x - \lambda^T(Ax - b) = (c^T - \lambda^T A)x + \lambda^T b.$$

So

$$(c^T - \lambda^T A)x = 0.$$

Also,

$$c^T x - \lambda^T(Ax - b) = c^T x$$

implies

$$\lambda^T(Ax - b) = 0.$$

On the other hand, suppose we have complementary slackness, i.e.

$$(c^T - \lambda^T A)x = 0 \text{ and } \lambda^T(Ax - b) = 0,$$

then

$$c^T x = c^T x - \lambda^T(Ax - b) = (c^T - \lambda^T A)x + \lambda^T b = \lambda^T b.$$

Hence by weak duality, x and λ are optimal. □

3.5 Simplex method

3.5.1 The simplex tableau

3.5.2 Using the Tableau

3.6 The two-phase simplex method

4 Non-cooperative games

4.1 Games and Solutions

Theorem (Nash, 1961)**.** Every bimatrix game has an equilibrium.

4.2 The minimax theorem

Theorem (von Neumann, 1928)**.** If $P \in \mathbb{R}^{m\times n}$. Then

$$\max_{x\in X}\min_{y\in Y} p(x,y) = \min_{y\in Y}\max_{x\in X} p(x,y).$$

Note that this is equivalent to

$$\max_{x\in X}\min_{y\in Y} p(x,y) = -\max_{y\in Y}\min_{x\in X} -p(x,y).$$

The left hand side is the worst payoff the row player can get if he employs the minimax strategy. The right hand side is the worst payoff the column player can get if he uses his minimax strategy.

The theorem then says that if both players employ the minimax strategy, then this is an equilibrium.

Proof. Recall that the optimal value of $\max\min p(x,y)$ is a solution to the linear program

$$\begin{aligned} &\text{maximize } v \text{ such that}\\ &\sum_{i=1}^m x_i p_{ij} \geq v \quad \text{for all } j = 1,\cdots,n\\ &\sum_{i=1}^m x_i = 1\\ &x \geq 0 \end{aligned}$$

Adding slack variable $z\in\mathbb{R}^n$ with $z\geq 0$, we obtain the Lagrangian

$$L(v,x,z,w,y) = v + \sum_{j=1}^n y_j\left(\sum_{i=1}^m x_i p_{ij} - z_j - v\right) - w\left(\sum_{i=1}^m x_i - 1\right),$$

where $w\in\mathbb{R}$ and $y\in\mathbb{R}^n$ are Lagrange multipliers. This is equal to

$$\left(1 - \sum_{j=1}^n y_j\right) v + \sum_{i=1}^m \left(\sum_{j=1}^n p_{ij}y_j - w\right) x_i - \sum_{j=1}^n y_j z_j + w.$$

This has finite minimum for all $v\in\mathbb{R}, x\geq 0$ iff $\sum y_i = 1$, $\sum p_{ij}y_j \leq w$ for all i, and $y\geq 0$. The dual is therefore

$$\begin{aligned} &\text{minimize } w \text{ subject to}\\ &\sum_{j=1}^n p_{ij}y_j \leq w \quad \text{for all } i\\ &\sum_{j=1}^n y_j = 1\\ &y\geq 0 \end{aligned}$$

This corresponds to the column player choosing a strategy (y_i) such that the expected payoff is bounded above by w.

The optimum value of the dual is $\min\limits_{y \in Y} \max\limits_{x \in X} p(x, y)$. So the result follows from strong duality. $\square$

Theorem. $(x, y) \in X \times Y$ is an equilibrium of the matrix game with payoff matrix P if and only if

$$\begin{aligned} \min_{y' \in Y} p(x, y') &= \max_{x' \in X} \min_{y' \in Y} p(x', y') \\ \max_{x' \in X} p(x', y) &= \min_{y' \in Y} \max_{x' \in X} p(x', u') \end{aligned}$$

i.e. the x, y are optimizers for the $\max \min$ and $\min \max$ functions.

5 Network problems

5.1 Definitions

5.2 Minimum-cost flow problem

5.3 The transportation problem

Theorem. Every minimum cost-flow problem with finite capacities or non-negative costs has an equivalent transportation problem.

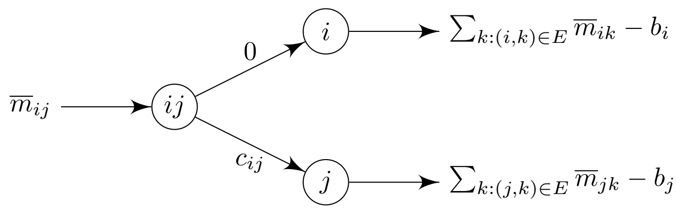
Proof. Consider a minimum-cost flow problem on network (V, E). It is wlog to assume that $\underline{m}_{ij} = 0$ for all $(i, j) \in E$. Otherwise, set $\underline{m}_{ij}$ to 0, $\overline{m}_{ij}$ to $\overline{m}_{ij} - \underline{m}_{ij}$, b_i to $b_i - \underline{m}_{ij}$, b_j to $b_j + \underline{m}_{ij}$, x_{ij} to $x_{ij} - \underline{m}_{ij}$. Intuitively, we just secretly ship the minimum amount without letting the network know.

Moreover, we can assume that all capacities are finite: if some edge has infinite capacity but non-negative cost, then setting the capacity to a large enough number, for example $\sum_{i\in V} |b_i|$ does not affect the optimal solutions. This is since cost is non-negative, and the optimal solution will not want shipping loops. So we will have at most $\sum |b_i|$ shipments.

We will construct an instance of the transportation problem as follows:

For every $i \in V$, add a consumer with demand $\left(\sum_{k:(i,k)\in E} \overline{m}_{ik}\right) - b_i$.

For every $(i, j) \in E$, add a supplier with supply $\overline{m}_{ij}$, an edge to consumer i with cost $c_{(ij,i)} = 0$ and an edge to consumer j with cost $c_{(ij,j)} = c_{ij}$.



The idea is that if the capacity of the edge (i, j) is, say, 5, in the original network, and we want to transport 3 along this edge, then in the new network, we send 3 units from ij to j, and 2 units to i.

The tricky part of the proof is to show that we have the same constraints in both graphs.

For any flow x in the original network, the corresponding flow on (ij, j) is x_{ij} and the flow on (ij, i) is $\overline{m}_{ij} - x_{ij}$. The total flow into i is then

$$\sum_{k:(i,k)\in E} (\overline{m}_{ik} - x_{ik}) + \sum_{k:(k,i)\in E} x_{ki}$$

This satisfies the constraints of the new network if and only if

$$\sum_{k:(i,k)\in E} (\overline{m}_{ik} - x_{ik}) + \sum_{k:(k,i)\in E} x_{ki} = \sum_{k:(i,k)\in E} \overline{m}_{ik} - b_i,$$

which is true if and only if

$$b_i + \sum_{k:(k,i)\in E} x_{ki} - \sum_{k:(i,k)\in E} x_{ik} = 0,$$

which is exactly the constraint for the node i in the original minimal-cost flow problem. So done. $\square$

5.4 The maximum flow problem

Theorem (Max-flow min-cut theorem)**.** Let δ be an optimal solution. Then

$$\delta = \min\{C(S) : S \subseteq V, 1 \in S, n \in V \setminus S\}$$

Proof. Consider any feasible flow vector x. Call a path $v_0, \cdots, v_k$ an *augmenting path* if the flow along the path can be increased. Formally, it is a path that satisfies

$$x_{v_{i-1}v_i} < C_{v_{i-1}v_i} \text{ or } x_{v_i v_{i-1}} > 0$$

for $i = 1, \cdots, k$. The first condition says that we have a forward edge where we have not hit the capacity, while the second condition says that we have a backwards edge with positive flow. If these conditions are satisfied, we can increase the flow of each edge (or decrease the backwards flow for backwards edge), and the total flow increases.

Now assume that x is optimal and let

$$S = \{1\} \cup \{i \in V : \text{ there exists an augmenting path from 1 to } i\}.$$

Since there is an augmenting path from 1 to S, we can increase flow from 1 to any vertex in S. So $n \notin S$ by optimality. So $n \in V \setminus S$.

We have previously shown that

$$\delta = f_x(S, V \setminus S) - f_x(V \setminus S, S).$$

We now claim that $f_x(V \setminus S, S) = 0$. If it is not 0, it means that there is a node $v \in V \setminus S$ such that there is flow from v to a vertex $u \in S$. Then we can add that edge to the augmenting path to u to obtain an augmenting path to v.

Also, we must have $f_x(S, V \setminus S) = C(S)$. Or else, we can still send more things to the other side so there is an augmenting path. So we have

$$\delta = C(S). \qquad \square$$