IB Optimisation: Lecture 5

Mike Tehranchi

University of Cambridge

4 May 2020

The barrier method. Given differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$, consider the problem to

$$P: \text{ minimise } f(x) \text{ subject to } g(x) \leq b$$

where f and g_i are convex for all i. (note: no regional constraint)

Now consider the family of unconstrained minimisation problems

$$P_\varepsilon : \ \text{minimise } f(x) - \varepsilon \sum_{i=1}^{m} \log(b_i - g_i(x))$$

(implicitly we have the constraint $g(x) < b$.)

Theorem (Convergence of the barrier method)

Suppose x^* is optimal for P and x_ε is optimal for P_ε. Then

$$0 \le f(x_\varepsilon) - f(x^*) \le m\varepsilon.$$

Duality preliminaries. The Lagrangian for P is

$$L(x, z, \lambda) = f(x) + \lambda^\top (b - z - g(x))$$

and the feasible Lagrange multipliers are

$$\Lambda = \{\lambda \in \mathbb{R}^m : \inf_{x \in \mathbb{R}^n, z \geq 0} L(x, z, \lambda) > -\infty\}$$

Claim: Let $\lambda \in \mathbb{R}^m$ be such that $\lambda \leq 0$ and such that there exists a $x_\lambda \in \mathbb{R}^n$ such that

$$Df(x_\lambda) = \sum_{i=1}^{m} \lambda_i Dg_i(x_\lambda).$$

then $\lambda \in \Lambda$ and the dual objective function is $h(\lambda) = b^\top \lambda + f(x_\lambda) - \lambda^\top g(x_\lambda)$.

[Recall that if $F : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable and $DF(\xi) = 0$, then ξ minimises F.]

Proof of claim. For $x \in \mathbb{R}^n$ and $z \geq 0$ we have the bound

$$\begin{aligned} L(x, z, \lambda) &= b^\top \lambda + f(x) - \lambda^\top g(x) - \lambda^\top z \\ &\geq b^\top \lambda + f(x_\lambda) - \lambda^\top g(x_\lambda) \end{aligned}$$

since $f(x) - \lambda^\top g(x)$ is convex and $\lambda^\top z \leq 0$. There is equality if $x = x_\lambda$, and the pair (z, λ) satisfy complementary slackness. □

Proof of convergence of the barrier method.
[Recall that if $F : X \to \mathbb{R}$ is differentiable, where $X \subseteq \mathbb{R}^n$ is open, and $\xi \in X$ minimises F, then $DF(\xi) = 0$.]
Since x_ε is optimal for P_ε we necessarily have

$$Df(x_\varepsilon) = -\varepsilon \sum_{i=1}^{m} \frac{Dg_i(x_\varepsilon)}{b_i - g_i(x_\varepsilon)}.$$

Let

$$\lambda_i = \frac{-\varepsilon}{b_i - g_i(x_\varepsilon)} \text{ for } 1 \leq i \leq m.$$

Note $\lambda < 0$ and

$$Df(x_\varepsilon) = \sum_{i=1}^{m} \lambda_i Dg_i(x_\varepsilon).$$

Hence $\lambda \in \Lambda$.

By the optimality of x^* and weak duality, we have

$$\begin{aligned} f(x_\varepsilon) \geq f(x^*) \geq h(\lambda) &= f(x_\varepsilon) + \lambda^\top (b - g(x_\varepsilon)) \\ &= f(x_\varepsilon) - m\varepsilon. \end{aligned}$$

□

Remark. Note since f and g_i are convex for all i, the value function

$$\varphi(b) = \inf\{f(x) : \quad g(x) \leq b\}$$

is convex. Fixing b and assuming there exists an optimiser x^* to problem P, by Lagrangian necessity, there exists λ^* such that

$$L(x^*, z^*) = \inf\{L(x, z, \lambda^*) : \quad x \in \mathbb{R}^n, z \geq 0\}$$

where $z^* = b - g(x^*)$.

Compare:

x^* optimal for P	x_ε optimal for P_ε
$\lambda^* \leq 0$	$\lambda_\varepsilon < 0$
$Df(x^*) = \sum_{i=1}^m \lambda_i^* Dg_i(x^*)$	$Df(x_\varepsilon) = \sum_{i=1}^m \lambda_{i,\varepsilon} Dg_i(x_\varepsilon)$
$(b - g(x^*))^\top \lambda^* = 0$	$(b - g(x_\varepsilon))^\top \lambda_\varepsilon = -m\varepsilon$

An algorithm

- Initial guess $x_0 \in \mathbb{R}^n$ such that $g(x_0) < b$.
- Initial $\varepsilon_0 > 0$
- For $k \geq 0$, solve P_{ε_k} approximately using your favourite algorithm (for instance, the gradient descent algorithm or Newton's method) starting from x_k.
- Let x_{k+1} be the approximate optimal solution.
- Let $\varepsilon_{k+1} = r\varepsilon_k$ where $0 < r < 1$.

Last time we saw that given the primal linear program

$$P: \text{ maximise } c^\top x \text{ subject to } Ax \leq b, \; x \geq 0,$$

the dual linear progam is

$$D: \text{ minimise } b^\top \lambda \text{ subject to } A^\top \lambda \geq c, \; \lambda \geq 0$$

Theorem (Fundamental theorem of linear programming, i.e. necessary and sufficient conditions for optimality)

Consider the problem to

$$P: \ \text{maximise } c^\top x \text{ subject to } Ax \le b, \ x \ge 0.$$

A vector $x^* \in \mathbb{R}^n$ *is optimal for problem* P *if and only if there exists a vector* $\lambda^* \in \mathbb{R}^m$ *such that*

- $Ax^* \le b, x^* \ge 0$ *(primal feasibility)*
- $A^\top \lambda^* \ge c, \lambda^* \ge 0$ *(dual feasibility)*
- $(b - Ax^*)^\top \lambda^* = 0 = (c - A^\top \lambda^*)^\top x^*$ *(complementary slackness)*

in which case λ^* *is optimal for the dual problem* D*, and the value of the two problems* $c^\top x^* = b^\top \lambda^*$ *agree.*

Proof of 'if'. Note that if x and λ are feasible for the respective problems, then by weak duality we have $c^\top x \le b^\top \lambda$. But if x^* and λ^* satisfy complementary slackness, then $c^\top x^* = b^\top \lambda^*$. This shows $c^\top x \le c^\top x^*$ for all feasible x and $b^\top \lambda \ge b^\top \lambda^*$ for all feasible λ, proving the optimality of x^* and λ^*. $\square$

Remark. The optimality of x^* is just an application of the Lagrangian sufficiency theorem as reformulated in Lecture 4.

Definition

A function $f : X \to \mathbb{R}$ is *concave* if the function $-f$ is convex.

Proof of 'only if'. Note the objective function $f(x) = c^\top x$ and the function defining the functional constraint $g(x) = Ax$ are linear. In particular, f is concave and g is convex. This means the value function of P is concave. By Lagrangian necessity, there exists an optimal Lagrange multiplier λ^*. $\square$

Motivation. Consider the problem to

$$\text{maximise } \psi(x) \text{ subject to } x \in X.$$

Now suppose that the set X is convex and the function ψ is convex. This means that if $x, y \in X$ and $0 < p < 1$ then

$$\begin{aligned} \psi(px + (1-p)y) &\leq p\psi(x) + (1-p)\psi(y) \\ &\leq \max\{\psi(x), \psi(y)\}. \end{aligned}$$

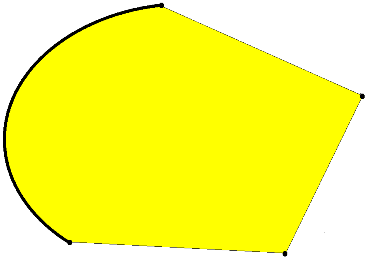
That is to say, the maximum of ψ on any segment $\{z : z = px + (1-p)y, 0 \leq p \leq 1\}$ occurs at one of the end points. Hence to find the maximum of ψ over X, we need only consider points of X that do not lie on a line segment contained in X.

Definition

Let $X \subseteq \mathbb{R}^n$ be a convex set. A point x is an *extreme point* if

$$x = py + (1-p)z$$

for $y, z \in X$ and $0 < p < 1$ implies $x = y = z$.



The extreme points of the convex set above are in bold.

For a linear maximisation program

- the objective function is concave, so we can characterise the optimal solution via duality
- the objective function is convex and the set of feasible solutions is convex, we can limit our search for optimal solutions to the extreme points of the set of feasible solutions.
- In particular, if the linear program has an optimal solution, then it must have an optimal solution that is an extreme point of the set of feasible solutions.