MOAC WORKSHEET OPTIMIZATION

Working through the following exercises you will glean a quick overview/review of a few essential ideas of optimization techniques that you will need in the MOAC *course. This worksheet does* not *take the place of a course in probability, and you should consult a good textbook for a more in-depth, rigorous treatment. Harder exercises start with the words "Can you. . . ".*

Exercise 1 Let y depend on $\mathbf{x}$ through a function F. Show that maximization of $y = F(\mathbf{x})$ is equivalent to minimization of $-F(\mathbf{x})$.

Exercise 2 Suppose $\mathbf{x}$ takes a finite number of different values. Outline a procudere to find the value of $\mathbf{x}$ for which $y = F(\mathbf{x})$ attains a minimum. (Hint: go through a list, and keep track of the lowest image value so far.)

Exercise 3 Show that the maximum of $y = -x^2$ is found at $x = 0$.

Exercise 4 Show that the minimum of $y = (x - \alpha)^{2/3}$ occurs at $x = \alpha$.

Exercise 5 Show that the maximum of $\exp\{-(x_1 - \alpha)^2(x_2 - \beta)^2\}$ is located at $(x_1, x_2) = (\alpha, \beta)$.

Exercise 6 Consider $F(x) = x/(1 + x)$. Show that the maximum of $F(x)$ equals $\alpha/(1 + \alpha)$ subject to the constraint $-1 < x \leq \alpha$.

Exercise 7 With $F(x) = x/(1+x)$, show that $1 > F(x)$ for $x \geq 0$. Also show that $2 > F(x)$ for $x \geq 0$.

Exercise 8 With $F(x) = x/(1 + x)$, show that there is a positive value for ε such that $2 - \varepsilon > F(x)$ for $x \geq 0$. By constrast, show that there is *no* positive value for ε such that $1 - \varepsilon > F(x)$ for $x \geq 0$.

§ 1 In exercise 7, you established that both 1 and 2 are upper bounds to $F(x)$ in the range of x given. In exercise 8 you showed that, moreover, 1 is a *least uppper bound*. A least upper bound is also known as a *supremum*.

Exercise 9 Can you define the term *infimum* (also known as *greatest lower bound*)?

Exercise 10 Consider, once more, $F(x) = x/(1 + x)$. Show that $F(x)$ attains has *no* maximum on the x-interval $(-1, \alpha)$ (where $\alpha > -1$), but that is does have a supremum (namely, $\alpha/(1 + \alpha)$) on this interval.

Exercise 11 Let $F(x) = \alpha + \beta x$ where α and β are two non-negative constants. Suppose that x is restricted to some set $\mathcal{X}$. Show that $F(x)$ has a supremum provided that $\mathcal{X}$ has both a infimum supremum and a finite supremum.

Exercise 12 Find the minimum of $y = x_1^2 + x_2^2$ subject to the constraint $x_2 - \alpha + \beta x_1 = 0$. (Hint: use the constraint to eliminate x_2 from the equation for y.)

§ 2 The technique of the last exercise, of reducing the dimension by substitution, can be difficult to apply if the constraint cannot be put in explicit form (that is, in the form of an equation of which either x_1 or x_2 is the subject). An alternative trick is based on increasing the dimension of the problem with the introduction of an auxiliary parameter λ. The aim is to maximize a function $y = F(x_1, x_2)$ subject to a constraint of the form $g(x_1, x_2) = 0$.

Exercise 13 Consider the family of functions $F(x_1, x_2) + \lambda g(x_1, x_2)$ parametrized by λ. Show that the functions in this family agree (i.e. map to the same y-values) exactly at those points (x_1, x_2) where the condition is satisfied.

§ 3 The function $F(x_1, x_2) + \lambda g(x_1, x_2)$, considered at a *fixed* value of λ, is just a another function of two variables, which we may denote by $H(x_1, x_2, \lambda)$, and can be optimized in the usual manner ('unconstrained'). Let $(\widehat{x}_{1\lambda}, \widehat{x}_{2\lambda})$ denote the location of this maximum.

Exercise 14 Show that, for every value of λ you may write

$$H(\widehat{x}_{1\lambda}, \widehat{x}_{2\lambda}) \geq F(x_1, x_2) \quad \text{for every } x_1, x_2 \text{ such that } g(x_1, x_2) = 0. \tag{1}$$

(Hint: use the fact that $H(\widehat{x}_{1\lambda}, \widehat{x}_{2\lambda}) \geq H(x_1, x_2, \lambda)$ everywhere, by definition of maximum, and your finding at exercise 13.)

Exercise 15 Suppose $\widehat{\lambda}$ is such that $g(\widehat{x}_{1\widehat{\lambda}}, \widehat{x}_{2\widehat{\lambda}}) = 0$. Show that $(\widehat{x}_{1\widehat{\lambda}}, \widehat{x}_{2\widehat{\lambda}})$ is in fact the constrained maximum you seek.

Exercise 16 Show that the maximum of the function $F(x_1, x_2)$, subject to the constraint $g(x_1, x_2) = 0$, if it exist, may be among the solutions of the system

$$\frac{\partial H(x_1, x_2, \lambda)}{\partial x_1} = 0$$
$$\frac{\partial H(x_1, x_2, \lambda)}{\partial x_2} = 0$$
$$\frac{\partial H(x_1, x_2, \lambda)}{\partial \lambda} = 0$$

where $H(x_1, x_2, \lambda) = F(x_1, x_2) + \lambda g(x_1, x_2)$. (Hint: the first two equations locate an unconstrained maximum for an arbitrary value of λ; the last one fixes λ by ensuring that the solution is on the locus of the constraint g.)

§ 4 The technique you have established in the last few exercises is called the technique of *Lagrange multipliers*; λ is an example of a Lagrange multiplier.

Exercise 17 Rework exercise 12 using a Lagrange multiplier.

Exercise 18 Generalize the technique to multiple Lagrange multipliers: outline the solution method for the optimization of $F(\mathbf{x})$ subject to $g_1(\mathbf{x}) = 0$ and $g_2(\mathbf{x}) = 0$, based on two Lagrange multipliers λ_1, λ_2.

§ 5 An important application of Lagrange multipliers is in optimal process control. The variables here are $x_1, \ldots, x_N$ and $u_1, \ldots, u_N$, for some integer N, where the objective is to maximize the following quantity:

$$\sum_{i=1}^{N} h_i(x_i, u_i)$$

with x_1 fixed at some given value. The variables x_i are interpreted as the values assumed by the state variable of a one-dimensional discrete-time dynamic system at times $t_1, \ldots, t_i, \ldots, t_N$.

Exercise 19 Verify that x_1 is the initial condition,

§ 6 Similarly, the u_i are the values of a forcing function at the successive instants in time $t_1, t_2, \ldots$. The constraint is that the variables must satisfy the system's state transition function:

$$x_{i+1} = F(x_i, u_i) \quad \text{for } i = 1, \ldots, N-1. \tag{2}$$

Exercise 20 Show that the latter constraint can be accommodated by introducing $N-1$ Lagrange multipliers, giving the objective function:

$$H = \sum_{i=1}^{N} h_i(x_i, u_i) + \sum_{i=1}^{N-1} \lambda_i \left(F(x_i, u_i) - x_{i+1} \right) .$$

§ 7 To determine an optimal solution, you differentiate H with respect to $x_2, \ldots, x_N$, $u_1, \ldots, u_N$ and $\lambda_1, \ldots, \lambda_{N-1}$ and you set all these derivatives equal to 0.

Exercise 21 How many equations to you obtain in this way?

Exercise 22 Verify that roughly a third of these equations just give you back the system's dynamics (i.e. the state transitions according to equation—(2)).

Exercise 23 Show that you also obtain $N-1$ equations of the form

$$\lambda_i = h^{\prime x}_{(}i+1)(x_{i+1}, u_{i+1}) + \lambda_{i+1} F^{\prime x}(x_{i+1}, u_{i+1}) \tag{3}$$

for $i = N-1, N-2, \ldots, 3, 2, 1$ where $\prime x$ indicates partial derivatives with respect to x (the first argument of these functions) and $\lambda_N = 0$ is a boundary condition.

Exercise 24 Interpret the Lagrange multiplier λ_i as the value assumed by a new state (the so-called *co-state*) at time t_i. What is its state transition function?

§ 8 Let

$$\pi_i(x_i, u_i) = h^{\prime u}_{(}i)(x_i, u_i) + \lambda_i F^{\prime u}(x_i, u_i)$$

where $\prime u$ indicates partial derivatives with respect to u (the second argument of these functions). Show that the final 'third' of equations is of the form

$$\pi_i(x_i, u_i) = 0 \ . \tag{4}$$

Exercise 25 Suppose that for all i, u_i is constrained to an interval $[u_{\min}, u_{\max}]$, and that there is no u-value in this interval for which equation (4). Can you show that H is maximized by putting $u_i = u_{\max}$ whenever $\pi_i(x_i, u_i) > 0$ in the allowed interval, and $u_i = u_{\min}$ whenever $\pi_i(x_i, u_i) < 0$?

§ 9 The last exercise makes clear why π is called the *switching function*. If the switching function assumes the value zero at some point in time t_i, the associated control $u = u_i$ at that instant is said to be *singular*.

Exercise 26 Can you extend the procedure to higher-dimensional discrete-time dynamical systems?

§ 10 The present approach to discrete-time optimal control problems can be applied to continuous time as well. In preparation for this, we define $h_i(x_i.u_i) = h(t, x_i, u_i)\Delta t$ and $f(x_i, u_i)\Delta t = F(x_i, u_i) - x_i$.

Exercise 27 Verify that $F'^x(x_i, u_i) = 1+f'^x(x_i, u_i)\Delta t$ and $F'^u(x_i, u_i) = f'^u(x_i, u_i)\Delta t$.

Exercise 28 Derive the following:

$$\frac{x_i - x_{i-1}}{\Delta t} = f(x_{i-1}, u_{i-1})$$

and

$$\frac{\lambda_i - \lambda_{i+1}}{\Delta t} = -\left(h'^x(t, x_i, u_i) + \lambda_i f'^x(x_i, u_i)\right) .$$

Exercise 29 Can you now establish the following? For a process $x(t)$ described by the differential equation $\dot{x} = f(x(t), u(t))$ with $x(t_1)$ given as initial condition, where $u(t)$ is the controlling (or 'input') function, an optimal control regime $u(\cdot)$ with regard to the objective functional

$$J = \int_{t_1}^{t_2} h(s, x(s), u(s))ds$$

may be sought by optimizing

$$H = h(t, x, u) + \lambda f(x, u) \tag{5}$$

with respect to u at each t, with $\lambda(t_2) = 0$, where

$$\dot{\lambda} = -\frac{\partial H}{\partial x} .$$

Exercise 30 With H given by equation (5), show that

$$\dot{x} = \frac{\partial H}{\partial \lambda} .$$

(Hint: you already know that $\dot{x} = f(x, u)$.)

Exercise 31 Suppose that $u(t)$ must satisfy $0 \leq u(t) \leq 1$ for all t, and the aim is to maximize J. Establish the following (candidate[1]) optimal control: $u(t) = 0$ where $\pi(t) < 0$, $u(t) = 1$ where $\pi(t) > 0$, with $\pi(t) = \frac{\partial H}{\partial u}$.

§ **11** If $\pi(t) = 0$ only at isolated points in time (the *switching moments*), control as stipulated by the last exercise is said to be *bang-bang*. On the other hand, if there is a time interval of positive duration, say $[t_a, t_b]$ where $t_b - t_a > 0$ and $t_1 \leq t_a < t_b \leq t_2$, such that $\pi(t) = 0$ for all $t \in [t_a, t_b]$, then control is *singular* between t_a and t_b (there may be none, one or more such intervals).

Exercise 32 Suppose optimal control is singular for all $t \in [t_1, t_2]$. Show that

$$\lambda(t) = -\frac{h_{\prime u}(t, x(t), u(t))}{f_{\prime u}(x(t), u(t))} \ .$$

(Hint: consider $\pi(t)$.)

Exercise 33 Suppose that h does not depend explicitly on t ($h_{\prime t} = 0$), and that optimal control is singular everywhere, as in the previous exercise. Can you show that $\dot{H} = 0$?

Exercise 34 With the assumptions of the previous exercise, show that (prospective) optimal control $u(t)$ is determined by the equation

$$h(x(t), u(t)) - \frac{h_{\prime u}(x(t), u(t)) f(x(t), u(t))}{f_{\prime u}(x(t), u(t))} = K \tag{6}$$

where K is a constant to be determined from the boundary condition $\lambda(t_2) = 0$.

§ **12** An equation such as (6), through which $u(t)$ can be found from $x(t)$ at every t (i.e. 'instantaneously') is said to give control in *feedback form*.

Exercise 35 Consider the problem where an extremum of the integral $\int_{t_1}^{t_2} h(t, x, \dot{x}) dt$ is sought. Show that the 'optimal path' $x(t)$ must satisfy

$$\frac{d}{dt}\frac{\partial h}{\partial \dot{x}} = \frac{\partial h}{\partial x} \ . \tag{7}$$

(Hint: consider $\dot{x} = u$ and assume that optimal control u is non-singular everywhere.)

§ **13** The problem of the last exercise is a simple example of the sort of problems addressed by the *calculus of variations*[2].

Exercise 36 Can you rewrite equation (7) as follows?

$$\frac{d}{dt}\left(h - \dot{x}\frac{\partial h}{\partial \dot{x}}\right) = \frac{\partial h}{\partial t} \ .$$

[1] We are not pursuing the subject with the rigour required to give assurances or detailed conditions; "insight" not "proof" is our watchword in these exercises.

[2] In the calculus of variations, equation (7) is known as *Euler's equation* or the *Euler-Lagrange equation*.

Exercise 37 Suppose that $h_{\prime t} = 0$ everywhere. Then show that $h - \dot{x}\frac{\partial h}{\partial \dot{x}}$ is equal to some (as yet unknown) constant. (Hint: the previous exercise.)

§ 14 The following exercises consider a planar objective function of the form

$$F(x_1, x_2) = \alpha + \beta x_1 + \gamma x_2$$

where α, β and γ are constants, the latter two non-zero; the objective function F is to be maximized subject to $(x_1, x_2) \in \mathcal{X}$ where $\mathcal{X}$ is a closed, simply connected subset of the (x_1, x_2)-plane.

Exercise 38 Consider a step $(\Delta x_1, \Delta x_2)$ in the (x_1, x_2)-plane. Verify that the associated change in objective value is given by $\Delta F = \beta \Delta x_1 + \gamma \Delta x_2$.

Exercise 39 Consider a *unit step* satisfying $(\Delta x_1)^2 + (\Delta x_2)^2 = 1$. If the step is at angle φ relative to the x_1-direction, show that $\Delta x_1 = \cos \varphi$ and $\Delta x_2 = \sin \varphi$.

Exercise 40 Show that an extremum for ΔF under a unit step of angle φ is obtained for the φ-values that solve

$$\tan \varphi = \frac{\gamma}{\beta}\,. \tag{8}$$

(Hint: differentiate with respect to φ.)

Exercise 41 Verify that equation (8) has *two* solutions, one of which indicates the direction of *steepest ascent*, the other of which the direction of *steepest descent*. (Hint: draw a graph of $\tan \varphi$ over a full period.)

Exercise 42 Consider the *gradient*

$$\nabla F = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$$

which you should realize is a *vector*. Show that ∇F points in the direction of steepest ascent, and that $-\nabla F$ points in the direction of steepest descent[3].

Exercise 43 Verify that the gradient is constant and non-zero throughout the domain $\mathcal{X}$. Hence conclude that the maximum exists and is unique.

§ 15 An *internal point* of the set $\mathcal{X}$ has the property that there is a positive ε such that all points within a distance less than ε are contained in $\mathcal{X}$.

Exercise 44 Confirm this definition with your intuitive notion of 'internal'.

Exercise 45 Consider a line segment PQ connecting at an arbitrary internal point P of the set $\mathcal{X}$ with a point Q on the boundary of $\mathcal{X}$, such that PQ points in the direction of ∇F. Show that $F(Q) > F(P)$. (Hint: the notation means what you would expect; if x_1^Q, x_2^Q are the coordinates of the point Q, then $F(Q) = F(x_1^Q, x_2^Q)$.)

[3] In fact, this result holds generally, not just for planar objective functions.

Exercise 46 Show that a maximum of F is to be found on the *boundary* of $\mathcal{X}$. (Hint: the previous exercise.)

Exercise 47 Let the relation $B(x_1, x_2) = 0$ define the boundary of $\mathcal{X}$. Show that an extremum of F on this boundary must satisfy the equation

$$\frac{B_{\prime x_1}(x_1, x_2)}{B_{\prime x_2}(x_1, x_2)} = \frac{\beta}{\gamma} \tag{9}$$

in addition to the condition $B = 0$. (Hint: use a Lagrange multiplier.)

Exercise 48 Suppose that the boundary of $\mathcal{X}$ is a polygon. Argue why, generically, you will find the maximum only at one of the corner points[4] of this polygon. (Hint: what does equation (9) look like for any given 'edge' of the polygon?)

Exercise 49 Suppose that the boundary of $\mathcal{X}$ is a polygon. Consider a vertex V and the edge VW connecting V to a neighbouring vertex W. Show that the component of the gradient *along* VW can only point toward V if the angle between the edge VW and the gradient is greater than $\frac{\pi}{2}$.

Exercise 50 Suppose that the boundary of $\mathcal{X}$ is a polygon. Consider a vertex V, its two neighbouring vertices W and U as well as the edges UV and VW. Suppose that the component of the gradient along UV points toward V, so that $F(V) > F(U)$, and that the component of the gradient along VW also points toward V, so that $F(V) > F(W)$. Consider the sum of the angles (i) between UV and the gradient and (ii) between Vw and the gradient; show that this sum exceeds π.

Exercise 51 Continuing the previous exercise, suppose that the maximum of F is achieved on some vertex M other than V ($M \neq V$). Since $F(V) > F(U)$ and $F(V) > F(W)$, you also have $M \neq U$ and $M \neq W$ (verify this). Consider the sum of the angles (i) between UV and MV and (ii) between Vw and MV; can you show that this sum, too, exceeds π?

§ 16 A simply connected domain is called *convex* if every line segment connecting two arbitrary points of the domain lies entirely within the domain.

Exercise 52 Show that a disc is convex.

Exercise 53 Draw the outline of a sea-star. Find two points such that the straight line segment connecting them lies only partly in your sea-star domain, thus establishing that the domain is not convex.

Exercise 54 In exercise 51 you considered the case where $\mathcal{M}$ has a polygonal boundary; the maximum is achieved on vertex M and there are three other vertices U, V and W such that $F(V) > F(U)$ and $F(V) > F(W)$. Show that $\mathcal{M}$ cannot be convex.

[4] A corner point is usually called a *vertex*, plural *vertices*.

Exercise 55 Suppose that $\mathcal{M}$ has a polygonal boundary and is convex. Let $F(M) > F(x)\ \forall x \in \mathcal{X}$. Show that M is a vertex and is, moreover, the only vertex V such that $F(V)$ is greater than both values of F at the two neighbouring vertices of V.

§ 17 The last exercise validates the following procedure to find the maximum of a planar objective function on a convex domain with polygonal boundary: start at any vertex; if it has a neighbour where F is greater, move toward that neighbour[5]; continue moving toward 'superior' neighbouring vertices until you are at a vertex both of whose neighbours have a smaller F-value.

Exercise 56 Consider a problem of the following appearance:

$$\begin{aligned} \text{to maximize} F(x_1, x_2) &= \alpha_0 + \beta_0 x_1 + \gamma_0 x_2 \\ \text{subject to} 0 &\leq \alpha_i + \beta_i x_1 + \gamma_i x_2 \end{aligned}$$

for $i = 1, \ldots, N$ where $N \geq 3$. All parameters are real and β_0, γ_0 are non-zero. Verify the following facts: (i) the constraints define a convex domain with a polygonal boundary; (ii) each vertex of the boundary defines a point where *two* of the N constraints are satisfied *with equality*; (iii) the problem is solved by applying the "vertex move" method outlined above, with vertices translated into pairs of equations as per (i) and (ii). (Hint: for (i), first establish that a convex domain is bisected by a straight line into two convex domains, then use induction; for (ii) observe that the constraints define the edges.)

Exercise 57 Is the converse of statement (ii) in the previous exercise true? (Hint: that is, does every pair of constraints, taken with equality, define a vertex on the boundary of the domain of x-values allowed by the set of constraints?)

§ 18 Essentially the same method can be applied to the n-dimensional version of the problem[6]:

$$\begin{aligned} \text{to maximize} F(x_1, \ldots, x_n) &= \kappa_{00} + \sum_{j=1}^{n} \kappa_{0j} x_j \\ \text{subject to} 0 &\leq \kappa_{i0} + \sum_{j=1}^{n} \kappa_{ij} x_j \end{aligned}$$

with $i = 1, \ldots, N$, $N \geq n + 1$, where the parameters κ_{ij} are real numbers and the κ_{0j}, $j = 1, \ldots, n$ are nonzero.

Exercise 58 Can you show that the domain is convex? (Hint: show that a straight line in $\mathbb{R}^n$ will intersect an $n-1$-hyperplane at most once, and use that fact.)

[5] Scanning the neighbours of your starting vertex, you may either move to the first 'superior' one that crops up, or you may scan both neighbours first and then, if both are superior, move to the 'best' of them.

[6] Known to operations researchers as a *linear program*; the somewhat startling term 'program' hails from the fact that problems of this sort often arise as the mathematical representation of a resource allocation task faced by an organization such as a company.

Exercise 59 Let $\eta_1, \ldots, \eta_n$ be n real parameters, not all zero, and use these to define a step Δx such that $\Delta x_j = \eta_j \kappa_{0j} / \sqrt{\sum_{j=1}^n \eta_j^2 \kappa_{0j}^2}$. Verify that Δx is a unit step in n-dimensional space.

Exercise 60 Show that the change in objective value associated with the unit step defined in the previous exercise is

$$\Delta F = \frac{\sum_{j=1}^n \eta_j \kappa_{0j}^2}{\sqrt{\sum_{j=1}^n \eta_j^2 \kappa_{0j}^2}} \ .$$

Exercise 61 Show that an extremum of the change ΔF calculated in the previous exercise is found when η_1 through η_n are all equal, say $\eta_j \equiv \eta$. Show that for $\eta > 0$, the unit step Δx defined in exercise 59 is in the direction of the gradient ∇F, and that for $\eta < 0$, Δx points in the direction of $-\nabla F$. (Hint: set $\frac{\partial F}{\partial \eta_j}$ equal to 0 for all j to find the extremum.)

Exercise 62 Can you extend the idea of exercise 48 to show that, generically, the maximum of F is found at a vertex of the domain? (Hint: each vertex is defined by n equalities, chosen out of the set of N constraints.)

§ 19 The natural extension of the search method is to consider vertices defined by n-tuples of equalities chosen out of the N given constraints. As before, a move from one vertex to its neighbour corresponds to the replacing one of the equations of the n-tuples with an equality taken from the remaining $N - n$ constraints.

Exercise 63 The crux to the validity of the "vertex move" method is again to prove that a vertex that is superior to all of its neighbours must be the unique maximum. Can you furnish this proof? (Hint: suppose, to the contrary, that there is a vertex V other than the maximum which 'dominates' all its neighbours. Show that there is an $n-1$ hyperplane $\mathcal{P}$ such that (i) $F(x) = F(V)$ for all x on that hyperplane; (ii)—the edges connecting V to its neigbours are all on one side of $\mathcal{P}$; and (iii) the maximum must be on the *other* side of $\mathcal{P}$. Then show that convexity enforces a contradiction.)