Chapter 2
Mathematics of Optimal Control

2.1. Optimization and optimal control models[1]

This chapter discusses mathematical ideas and techniques relevant to optimization questions in economics and related areas, and particularly relevant for construction and application of social choice models, based on the assumptions of new[3] economics discussed in chapter 1. First considered is a *static* model, optimising over a vector variable z, typically with a finite number of components. When the variable z is not static, but describes some variation over time, an optimal control model may be required, where the objective is typically an integral over a time horizon, say $[0,T]$), with perhaps an additional term at the final time T, and the evolution over time is described by a dynamic equation, typically a differential equation. This leads to an optimal control model, where z becomes a function of time t. In each case, a minimum may be described by necessary *Karush-Kuhn-Tucker (KKT) conditions*, involving Lagrange multipliers. When the time t is a continuous variable, a related set of necessary *Pontryagin conditions* often apply (see section 2.2).

There are also discrete-time models, where the integration is replaced by summation over a discrete time variable, say $t = 0, 1, 2, ..., T$, and the dynamic equation is a difference equation. For discrete time, the KKT conditions apply, but not all the Pontryagin theory.

Questions arise of *existence* (thus, when is a maximum or minimum reached?), *uniqueness* (when is there exactly one optimum?), relaxation of the usual assumption of convex functions, and what happens to a dual problem (in which the variables are the Lagrange multipliers) in the absence of convex assumptions? These are discussed in sections 2.3 through 2.7. Further issues arise when there are several conflicting objectives; these are discussed in sections 2.8 through 2.14.

Consider first a mathematical programming model (e.g a model for a normative social choice problem in economics and finance):

$$\text{MIN } f(z) \text{ subject to } g(x) \leq 0, k(z) = 0,$$

in which an objective function $f(\cdot)$ is maximized, with the state variable z constrained by inequality and equality constraints. The functions f, g and h are assumed differentiable. Note that a maximization problem, MAX $f(z)$, may be considered as minimization by MIN $-f(z)$. Assume that a local minimum is reached at a point $z = p$ (*local* means that $f(z)$ reaches a minimum in some region around p, but not necessarily over all values of z satisfying the

[1] See also sections 4.2, 4.5, 6.3, 9.4.

constraints.) Assume that z has n components, $g(z)$ has m components, and $k(z)$ has r components. The gradients $f'(p),\ g'(p),\ k'(p)$ are respectively $1 \times n,\ m \times n,\ r \times n$ matrices.

Define the *Lagrangian* $L(z) := f(a) + \rho g(z) + \sigma k(z)$. The Karush-Kuhn-Tucker necessary conditions (KKT):

$$L'(p) = 0,\ \ \rho \geq 0,\ \ \rho g(p) = 0$$

then hold at the minimum point p, for some Lagrange multipliers ρ and σ, provided that some *constraint qualification* holds, to ensure that the boundary of the *feasible region* (satisfying the constraints) does not behave too badly). The multipliers are written as row vectors, with respectively m and r components. These necessary KKT conditions are *not* generally sufficient for a minimum. In order for (KKT) at a feasible point p to imply a minimum, some further requirement on the functions must be fulfilled. It is enough if f and g are convex functions, and k is linear. Less restrictively, *invex* functions may be assumed - see section 2.4.

Consider now an optimal control problem, of the form:

$$\mathrm{MIN}_{x(.),u(.)}\ F^0(x,u) := \int_0^1 f(x(t),u(t),t)dt + \Phi(x(1))$$

$$\text{subject to } x(0) = a,\ \dot{x}(r) = m(x(t),u(t),t),\ q(t) \leq u(t) \leq r(t)\ \ 0 \leq t \leq 1).$$

(This problem can represent the problem of optimizimg the intertemporal welfare in an economics or finance model.) Here the state function $x(.)$, assumed piecewise smooth, and the control function $u(.)$, assumed piecewise continuous, are, in general, vector-valued; the inequalities are pointwise. A substantial class of optimal control problems can (see Craven, 199/5); Craven, de Haas and Wettenhall, 1998) be put into this form; and, in many cases, the control function can be sufficiently approximated by a step-function. A terminal constraint $\sigma(x(1)) = b$ can be handled by replacing it by a penalty term added to $F^0(x,u)$; thus the objective becomes:

$$F(x,u) := F^0(x,u) + \tfrac{1}{2}\mu\|\sigma(x(1)) - b^*\|^2,$$

where μ is a positive parameter, and k approximates to b. In the augmented Lagrangian algorithm (see e.g. Craven, 1978), constraints are thus replaced by penalty terms; μ is finite, and typically need not be large; here $b^* = b + \theta/\mu$, where θ is a Lagrange multiplier. If there are few constraints (or one, as here), the problem may be considered as one of parametric optimization, varying b^*, without computing the multipliers. Here T is finite and fixed; the endpoint constraint $q(x(T)) = 0$ is not always present; constraints on the control $u(t)$ are not always explicitly stated, although an implicit constraint $u(t) \geq 0$ is commonly assumed. If $q(.)$ or $\Phi(.)$ are absent from the model, they are replaced by zero.

In a model for some economic or financial question of maximizing welfare, the state $x(.)$ commonly describes capital accumulation, and the control $u(.)$ commonly describes consumption. Both are often vector functions.

The differential equation, with initial condition, determines $x(.)$ from $u(.)$; denote this by $x(t) = Q(u)(t)$; then the objective becomes:

$$J(u) = F^0(Q(u), u) + \tfrac{1}{2}\mu\|\sigma(Q(u)(1)) - b^*\|^2,$$

Necessary Pontryagin conditions for a minimum of this model have been derived in many ways. In Craven (1995), the control problem is reformulated in mathematical programming form, in terms of a Lagrangian:

$$\int_0^T [e^{-\delta t} f(x(t), u(t)) + \lambda(t)m(x(t), u(t), t) - \lambda(t)\dot{x}(t) + \alpha(t)(q(t) - u(t))$$

$$+\beta(t)(u(t) - r(t) + \frac{1}{2}\mu[\Phi(x(t) - \mu^{-1}\rho]_+^2 + \frac{1}{2}\mu[q(x(T) - \mu^{-1}\nu]\delta(t - T)]\ dt.$$

with the costate $\lambda(t)$, and also $\alpha(t)$ and $\beta(t)$, representing Lagrange multipliers, μ a weighting constant, ρ and ν are Lagrange multipliers, and $\delta(t-T)$ is a Dirac delta-function. Here, the terminal constraint on the state, and the endpoint term $\Phi(x(T))$ in the objective, have been replaced by penalty cost terms in the integrand; the multipliers ρ and ν have meanings as shadow costs. (This has also computational significance $-$ see section 3.1. The solution of a two-point boundary value problem, when $x(T)$ is constrained, has been replaced by a minimization.) The state and control functions must be in suitable spaces of functions. Often $u(.)$ is assumed piecewise continuous (thus, continuous, except for a finite number of jumps), and $x(.)$ is assumed piecewise smooth (the integral of a piecewise continuous function.)

The adjoint differential equation is obtained in the form:

$$-\dot{\lambda}(t) = e^{-\delta t} f_x(x(t), u(t)) + \lambda(t)m_x(x(t), u(t), t),$$

where f_x and m_x denote partial derivatives with respect to $x(t)$, together with a boundary condition (see Craven, 1995):

$$\lambda(T) = \Phi_x(x(T)) + \kappa q_x(x(T)),$$

in which Φ_x and q_x denote derivatives with respect to $x(T)$, and κ is a Lagrange multiplier, representing a shadow cost attached to the constraint $q(x(T)) = 0$. The value of κ is determined by the constraint that $q(x(T) = 0$. If $x(T)$ is *free*, thus with no terminal condition, and Φ is absent, then the boundary condition is $\lambda(T) = 0$. Note that $x(T)$ may be partly specified, e.g. by a linear constraint $\sigma^T x(T) = b$ (or $\geq$ b), describing perhaps an aggregated requirement for several kinds of capital. In that case, the terminal constraint differs from $\lambda(T) = 0$.

A diversity of terminal conditions for $\lambda(T)$ have been given in the economics literature (e.g. Sethi and Thompson, 2000); they are particular cases of the formula given above. For the constraint $q(x(T)) \geq 0$, the multiplier $\kappa \geq 0$.

From the standard theory, the gradient $J'(u)$ is given by:

$$J'(u)z = \int_0^1 (f + \lambda(t)m)_u(x(t), u(t), t)dt,$$

where the costate $\lambda(.)$ satisfies the adjoint differential equation:

$$-\dot{\lambda}(t) = (f + \lambda(t)m)_x(x(t), u(t), t),\ \ \lambda(1) = \mu(\sigma(x(1)) - b^*) + \Phi'(x(1));$$

$(.)_x$ denotes partial derivative. A constraint such as $\int_0^1 \theta(u(t))dt \leq 0$, which involves controls at different times, can be handled by adjoining an additional state component $y_0(.)$, satisfying $y_0(0) = 0,\ \dot{y}_0(t) = \theta(u(t))$, and imposing the state constraint $y_0(1) \leq 0$.The latter generates a penalty term $\frac{1}{2}\mu\|y_0(1)) - c\|_+^2$, where $c \approx 0$ and $[.]_+$ replaces negative components by zeros. (See an example of such a model in section 1.4).

2.2. Outline of the Pontryagin theory[2]

This section gives an outline derivation of the Pontryagin necessary conditions for an optimal control problem in continuous time, with a finite time horizon T. In particular, it is indicated how the boundary conditions for the costate arise; they are critical in applications. Comments at the end of this section i indicate what may happen with there is an *infinite time horizon* ($T = \infty$). That discussion is continued in Chapter 9.

Consider now an optimal control problem:

$$\text{MIN } F(z, u) := \int_0^T f(x(t), u(t), t)dt \text{ subject to}$$

$$\dot{x}(t) = m(x(t), u(t), t),\ \ x(0) = x_0,\ \ g(u(t), t) \leq 0\ \ (0 \leq t \leq T).$$

Here the vector variable z is replaced by a pair of functions, the *state function* (or *trajectory*) $x(\cdot)$ and the *control function* $u(\cdot)$. This problem can be written formally as:

$$\text{MIN } F(x, u) \text{ subject to } Dx = M(x, u),\ \ G(u) \leq 0.$$

Here D maps the trajectory $x(.)$ onto its gradient (thus the whole graph of $\dot{x}(t)$ $(0 \leq t \leq T)$. Consider the *Lagrangian* function :

$$L(x, u; \theta, \zeta) := F + \theta(-Dx + M) + \zeta G =$$

[2] See also sections 4.2, 5.7, 6.3.3, 6.5.2.

$$\int_0^T (f + \lambda m)dt - \int_0^T \lambda(t)\dot{x}(t)dt + \int_0^T \mu(t)g(u(t), t)dt,$$

in which the multiplier θ is represented by a *costate function* $\lambda(t)$, described by $\theta w = \int_0^T \lambda(t)w(t)dt$ for each continuous function w, and ζ is similarly represented by a function $\mu(.)$. Note that the *Hamiltonian* function:

$$h(x(t), u(t), t, \lambda(t)) := f(x(t), u(t), t) + \lambda(t)m(x(t)u(t), t)$$

occurs in the integrand. An integration by parts replaces the second integral by:

$$\lambda(T)x(T) - \lambda(0)x(0) + \int_0^T \dot{\lambda}(t)x(t),$$

in which $\lambda(0)x(0)$ may be disregarded, because of the initial condition on the state, $x(0) = x_0$.

If the control problem reaches a minimum, and certain regularity restrictions are satisfied, then necessary KKT conditions also hold for this problem, namely:

$$L_x = 0,\ L_u = 0, \zeta \geq 0,\ \zeta G = 0,$$

where suffixes x and u denote partial derivatives. The following is an outline of how the Pontryagin theory can be deduced, using (KKT). For a detailed account of this approach, especially including the (serious) assumptions required for its validity, see e.g. Craven (1995).

From $L_x = 0$ in (KKT), the *adjoint differential equation*:

$$-\dot{\lambda}(t) = h_x(x(t),\ \lambda(T)x(T) = 0$$

may be deduced, using the endpoint boundary condition $\lambda(T)x(T) = 0$ to eliminate the integrated part. The rest of (KKT) gives necessary conditions for minimization of the Hamiltonian with respect to the control only, subject to the constraints on the control. While they do not generally imply a minimum, they do in restrictive circumstances, leading to *Pontryagin's principle*, which states that the optimal control minimizes the Hamiltonian with respect to the control $u(t)$, subject to the given constraints on the control, while holding the state $x(t)$ and costate $\lambda(t)$ at their optimal values. The restrictions include the following:

- The control problem reaches a local minimum, with respect to the norm

$$\|u\|_1 := \int_0^T |u(t)|dt.$$

- The constraints on the control hold for each time t separately (so that constraints involving a combination of two or more times are excluded).
- Existence and boundedness of first and second derivatives of f and m.

The necessary Pontryagin conditions for a minimum of the control problem hence comprise:

- The dynamic equation for the state, with initial condition.
- The adjoint equation for the costate, with terminal condition.
- The Pontryagin principle.

If a terminal condition is omitted, then the system is not definitely defined, and generally uniqueness is lost. If the state has r components, then r terminal conditions are required. If $x^i(T)$ is not specified, then $\lambda^i(T) = 0$; if, however, $x^i(T)$ is fixed, then $\lambda^i(T)$ is free (not specified.)

This discussion has assumed a fixed finite time horizon T. If an infinite horizon is required, as in some economic models, then serious difficulties arise. It is not obvious that any minimum is reached. The objective $F(x, u)$ may be infinite, unless the function f includes a discount factor, such as $e^{-\delta t}$. The conditions on derivatives are not generally satisfied, over an infinite time domain, unless the state and control are assumed to converge sufficiently fast to limits as $t \to \infty$. The conjectured boundary condition $\lim_{t\to\infty} \lambda(t)x(t) = 0$ does not necessarily hold. Some circumstances where this boundary condition does hold are analysed in Chapter 9, with assumptions on convergence rates.

If the control problem is truncated to a finite planning interval $[0, T]$, with a terminal condition fixing $x(T)$ at the assumed optimal value for the infinite-horizon problem, then this gives the necessary conditions for the infinite-horizon problem, *except* that the terminal condition for the costate is omitted. So the system of conditions is not definitely defined, and often allows some additional, though spurious, solution. Various authors have adjoined a boundary condition (called *transversality condition* arbitrarily, to exclude the additional solutions. But it is preferable to obtain the correct boundary condition from a complete set of necessary conditions for a minimum (see Craven (2003) and Chapter 9.)

2.3. When is an optimum reached?

Questions arise of (i) existence of a maximum point $\mathbf{x}^*$, (ii) necessary conditions for a minimum, (iii) sufficient conditions for a maximum, (iv) uniqueness, (v) descriptions by *dual variables* (which interpret *Lagrange multipliers* as *prices*).

Consider first the maximization of an objective $f(x)$ over $x \in \mathbf{R}^n$. Concerning *existence*, if x is in $\mathbf{R}^n$, f and each g_i are continuous functions, and if the *feasible set* E of those x satisfying the constraints $g_i(x) \geq 0$ is *compact*, then at least one maximum point x* exists.

However, in an optimal control model in continuous time, the compactness property is usually not available. (If there are only a finite number of variables, then a set is compact if it is closed and bounded; but that does not hold for an infinite number of variables, as for example for a continuous state function.) Sometimes convex, or invex, assumptions can be used to show that an optimum is reached - see section 2.4.

Assuming that a maximum is reached for $f(x)$, subject to the inequality constraint $g(x) \geq 0$, then *necessary conditions* for p to be a minimum are that a Lagrange multiplier vector $\rho^* \geq \mathbf{0}$ exists, so that the Lagrangian $L(x, \mu) := f(x) + \rho g(x)$ satisfies the *Karush-Kuhn-Tucker conditions* (KKT) or the *saddlepoint condition* (SP):

$$\text{(KKT):} \qquad L_x(p, \rho*) = 0, \rho^* \geq 0, \quad \rho^* g(p) = 0, \ g(p) \geq 0;$$

$$\text{(SP):} \qquad L(\mathbf{p}, \rho) \geq L(p, \mu^*) \geq L(x, \rho^*) \text{ for all x, and all } \rho \geq 0.$$

The conditions often assumed for (KKT) are that f and each component g_i are differentiable functions. If there is also an equality constraint $k(x) = 0$, then a term $\sigma k(x)$ is added to the Lagrangian, and a regularity assumption is required, e.g. that the gradients of the active constraints (those $g_i'(p)$ for which $g_i p) = 0$, together with all the $h_j'(p)$) are linearly independent vectors. The conditions often assumed for (SP) (with k absent) are that f and each g_i are convex functions (which need not be differentiable), together with *Slater's condition,* that $g(z) > 0$ for some feasible point z. Then (SP) is a *sufficient condition* for a minimum (even without assuming convexity). However, (KKT) is not sufficient for a maximum; (KKT) implies a maximum if also f and each g_i are convex functions; this maximum point is unique if f is strictly convex.

If the functions f, g, k also contain a parameter q, then the optimal value of $f(p)$ also depends on q; denote this function by $V(q)$. Under some regularity conditions (see e.g. Fiacco and McCormick, 1968; Craven, 1995), the gradient $V'(\mathbf{0})$ equals the gradient $L_q(p, \rho^*, \sigma^*)$.

For a maximization problem, the inequalities for L in (SP) are reversed, and convexity applies to $-f$ and $-g$.

For the problem with f and g convex functions, and k a linear function, there is associated a *dual problem*:

$$\text{MAX } f(y) + vg(y) + wk(y) \text{ subject to } v \geq 0, f'(y) + vg'(y) + wk'(y) = 0.$$

Assume that the given problem reaches a minimum at $\mathbf{p}$, and that KKT holds with Lagrange multipliers ρ and σ. Then two properties relate the given *primal* problem and the dual problem:

- *Weak Duality* If x satisfies the constraints of the primal problem, and y, v, w satisfy the constraints of the dual problem, then:

$$f(x) \geq f(y) + vg(y) + wh(y);$$

- *Zero Duality Gap (ZDG)* The dual problem reaches a maximum when $(y, v, w) = (\mathbf{p}, \rho, \sigma)$. Thus, the Lagrange multipliers are themselves the solutions of an optimization problem.

These duality properties are well known for convex problems. However, they also hold (see Craven, 1995) when (f, g, k) satisfies a weaker, *invex* property, described in section 2.4. This property holds for some economic and finance models, which are not convex.

2.4. Relaxing the convex assumptions[3]

Convex assumptions are often not satisfied in real-world economic models (see Arrow and Intriligator, 1985). It can happen, however, that a global maximum is known to exist, and there is a unique KKT point; then that KKT point must be the maximum. This happens for a considerable class of economic models − see section 2.5. Otherwise, the necessary KKT conditions imply a maximum under some weaker conditions than convexity. It suffices if f is pseudoconvex, and each g_j is pseudoconcave, or less restrictively if the vector $-(f, g_1, \ldots, g_m)$ is *invex* . (Here k is assumed absent or linear).

A vector function h is *invex* at the point p (see Hanson, 1980; Craven, 1995) if, for some *scale function* η :

$$h(x) - h(p) \geq h'(p)\eta(x, p).$$

(The $\geq$ is replaced by $=$ for a component of k corresponding to an equality constraint.) Note that h is convex at p if $\eta(x, p) = x - p$; but *invex* occurs in other cases as well.

From Hanson (1980), a KKT point p is a minimum, provided that $(f, g_1, \ldots, g_m)$ is invex at p.

If inactive constraints are omitted then (Craven, 2002) this invex property holds exactly when the Lagrangian $L(x, \mu) = f(x) + \mu^T g(x)$ satisfies the *saddlepoint condition* $L(p, \mu) \geq L(p, \mu^*) \geq L(z, \mu^*)$ for all z and all $\mu \geq 0$. If the problem is transformed by $x = \varphi(y)$, where φ is invertible and differentiable, then h is invex exactly when $h_0\varphi$ is invex (with a different scale function). (Thus, the invex property is invariant to such transformations φ; then name *invex* derives from *invariant convex* .) If a φ can be found such that $h_0\varphi$ is convex, then it follows that h is invex. (This happens e.g, for the Kendrick-Taylor growth model - see Islam and Craven, 2001a).

For the problem: MAX $f(x)$ subject to $g(x) \geq 0$, where the vector function $(-f, -g)$ is assumed invex, a local minimum is a global minimum; if there are several local minima, they have the same values of $f(x)$. Under the further assumption of *strict invexity* for f at a minimum point p where (KKT) holds (with multiplier μ^*), namely that $f(x) - f(p) > f'(p)\eta(x, p)$ whenever $x \neq p$, a minimum point p is unique. For, with $L(x) = f(x) + \mu^* g(x)$, $L(.)$ is then strictly invex, hence:

$$f(x) - f(p) \geq L(x) - L(p) > L'(p)\eta(x, p) = 0.$$

For optimal control, as in section 2.2, the spaces are infinite dimensional; however KKT conditions (equivalent to Pontryagin, conditions), *quasimax* (see section 2.7), invexity, and related results apply without change. However, *invexity* may be difficult to verify for a control model, because the dynamic equation for $\dot{x}(t)$ is an equality constraint; thus convexity assumptions would require the dynamic equation to be linear.

[3] See also sections 4.4 and 8.6.

Invexity can sometimes be established for a control problem by a suitable transformation of the functions, assuming however that constraints on the control function are not active. Consider a model for *optimal growth*, (see Intriligator, 1971; Chiang, 1992), in which $u(t) = c(t)$ is consumption, $x(t) = k(t)$ is capital, $f(x(t), u(t)) = U(c(t))$ where the social welfare function $U(.)$ is positive increasing concave, and:

$$m(x(t), u(t), t) = \varphi(k(t)) - \text{ c(t) } - rk(t).$$

Consider now a dynamic equation $\dot{k}(t) = b(t)\theta(k(t)) - c(t)$, where $b(t) > 0$ is given . The transformation $x(t) = \psi(k(t))$ leads to the differential equation:

$$\dot{x}(t) = \psi'(\psi^{-1}(x(t))[b\varphi(\psi^{-1}(x(t)) - c(t))].$$

This assumes that the function $\psi(.)$ is strictly increasing, hence invertible, as well as differentiable. If ψ can be chosen so that $\psi'(.)\varphi(.) = 1$, then the differential equation becomes: $\dot{x}(t) = b(t) - u(t)$, where $u(t) := \psi'(\psi^{-1}(x(t))c(t)$ is a new control function. Then one may ask whether the integrand:

$$-e^{-rt}U(u(t)/\psi'(\psi^{-1}(x(t))$$

happens to be a *convex* function of $(x(t), u(t))$? If it is, then the problem was *invex*, since it could be transformed into a convex problem.

A similar approach was followed by Islam and Craven (2001a) for the Kendrick-Taylor model, defined in section 1.4. Here $\dot{x}(t) = \zeta e^{qt}k(t)^{\beta} - \sigma k(t) - c(t)$ and the integrand is $c(t)^{\tau}$, with $\beta = 0.6$ and $\tau = 0.1$. The transformation $x(t) = (k(t)e^{\sigma t})^{1-\beta}$, followed by $c(t) = x(t)^{\theta}u(t)$ for suitable θ, reduces the dynamic equation to the form:

$$\dot{x}(t) = (1-\beta)\zeta e^{rt} - (1-\beta)e^{\sigma t}u(t).$$

The integrand of the objective function becomes a function of $(x(t), u(t))$, which is concave if its matrix of second derivatives is negative definite. This holds for the given values of β and τ.

The approach of the previous paragraph seems to work a little more generally. However, there are difficulties if $x(t)$ has more than one component; and it must be assumed that any constraints on the control, such the constraint $0 \leq c(t) \leq \varphi(k(t))$ in Chiang (1992), are not active, since these will not transform to anything tractable.

The invex property can sometimes be used to establish existence.

$$\text{MIN } f(x) \text{ subject to } g(x) \leq 0, k(x) = 0,$$

satisfies convex, or invex, assumptions on the functions f, g, h. Without assuming that a minimum is reached, suppose that the necessary KKT conditions at a point $\mathbf{p}$ satisfying the constraints, with some Lagrange multipliers ρ and σ, can be solved for p, ρ, and σ. Define the Lagrangian: $L(x) := f(x) + \rho g(x) + \sigma k(x)$. If x satisfies the constraints, then:

$$f(x) - f(\mathbf{p}) \geq L(x) - L(\mathbf{p} \geq L'(\mathbf{p})\eta(x, \mathbf{p}) = 0.$$

Thus the problem reaches a global minimum at $\mathbf{p}$. This approach is not restricted to a finite number of variables, so it may be applied to optimal control.

It is well known (Mangasarian, 1969) that KKT conditions remain sufficient for an optimum if the convexity assumptions are weakened to $-f$ *pseudoconvex* x and each $-g_i$ *quasiconvex*. The definitions are as follows. The function f is *quasiconcave* at p if $f(x) \geq f(p) => f'(p)(x-p) \geq 0$, *pseudoncave* at p if $f(x) > f(p) => f'(p)(x-p) > 0$. If f is pseudoconcave, and each g_i is quasiconcave, then a KKT point is a maximum. If the function $n(.)$ is concave and positive, and $d(.)$ is convex and positive, then the ratio $n(.)/d(.)$ is pseudoconcave. Apart from this case (*fractional programming*), quasi- and pseudo-concave are more often assumed than verified. These assumptions can be further weakened as follows. The function h is *quasiinvex* at p if $h(x) \geq h(p) => h'(p)\eta(x,p) \geq 0$, *pseudoinvex* at p if $h(x) > h(p) => h'(p)\eta(x,p) > 0$. As above, the scale function η must be the same for all the functions describing the problem. Then $-f$ *pseudoinvex* and each $-g_i$ *quasiinvex* also make a KKT point a maximum. If $n(.) > 0$ is concave, and $d(.) > 0$ is convex, then a transformed function $-n \circ \varphi(.)/d \circ \varphi(.)$ is pseudoinvex, when φ is differentiable and invertible.

2.5. Can there be several optima?[4]

Many nonlinear programs have local optima that are not global. The evolution in time, in an optimal control model, puts some restrictions on what may happen with such a problem. The following discussion, from Islam and Craven (2004), gives conditions when an optimum is unique, or otherwise. Two classes of control problem often occur:

(a) When f and m are linear in $u(.)$, then *bang-bang control* often occurs, with $u(.)$ switching between its bounds, plus perhaps a singular arc. The optimum is essentially defined by the switching times; and the bounds on $u(.)$ are needed, for an optimum to be reached. Some such problems have several local optima, with different numbers of switching times.

(b) When f and m are nonlinear, and the controls on the constraints are inactive, then the Pontryagin necessary conditions for an optimum requires that:

(i) As well as the dynamic equation, the costate $\lambda\ (.)$ satisfies the differential equation :

$$-\dot{\lambda}(t) = e^{-\rho t} f_x(x(t), u(t)) + \lambda(t) m_x(x(t), u(t), t), \quad \lambda(T) = -\Phi'(x(T)),$$

[4] See also section 5.6.

or, for the Kendrick-Taylor example (see section 2.4):

$$-\dot{\lambda}(t) = e^{-\rho t}U'(c(t)) + \lambda(t)b(t)\theta'(k(t)),\ \ \lambda(T) = -\Phi'(k(T));$$

setting $z(t) = (k(t), \lambda(t))$, the two differential equations combine into one of the form:

$$\dot{z}(t) = \zeta(z(t), u(t), t),\ \ z(0) = z_0;$$

(the λ part of z_0 is a parameter, varied to satisfy the $\lambda(T)$ condition).

(ii) From Pontryagin's principle,

$$e^{-\rho t}f(x(t), u(t)) + \lambda(t)Pm(x(t), u(t), t)$$

is maximized over $u(t)$; or, for the example, $e^{-\rho t}U(c(t)) - \lambda(t)c(t)$ is maximized over $c(t)$ at the optimal $c(t)$; for this (in the absence of constraints on $c(t)$) it is necessary (but not sufficient) that the gradient $e^{-\rho t}U'(c(t)) - \lambda(t) = 0$.

If (I) the gradient equation in (ii) can be solved uniquely, and globally, for $u(t)$, say as

$$u(t) = p(\lambda(t), t),$$

then substitution into the $\dot{z}(t)$ equation gives n

$$\dot{z}(t) = Z(z(t), t) \equiv \zeta(z(t), p(\lambda(t), t), t), z(0) = z_0.$$

Assume additionally (II) that $Z(., t)$ satisfies a Lipschitz condition, uniformly in t. Then the differential equation has a unique solution $z(t)$. In these circumstances, the control problem has a unique optimum. (Note that (I) excludes any jumps in $u(.)$.)

Assumption (I) holds if the utility $U.)$ is concave and strictly increasing. In other cases, there may be several solutions for $u(t)$. In Figures 1 and 2, $U(.)$ is *quasiconcave* , and if $U'(u)$ lies in a certain range there are three solutions for u. However, the optimum control $u(t)$ may still be unique. A simple example, for a pseudoconcave objective (which implies quasiconcave) and a linear differential equation, is given in section 2.8.

In Kurz (1968), a class of optimal control models for economic growth are analysed using the Pontryagin theory. There is a unique optimum if certain concavity conditions are fulfilled. If they are not, then multiple optima may occur. Kurz gives numerical examples of multiple optima when the welfare function $U(.)$ depends on the state $k(.)$ as well as the control $u(.)$.

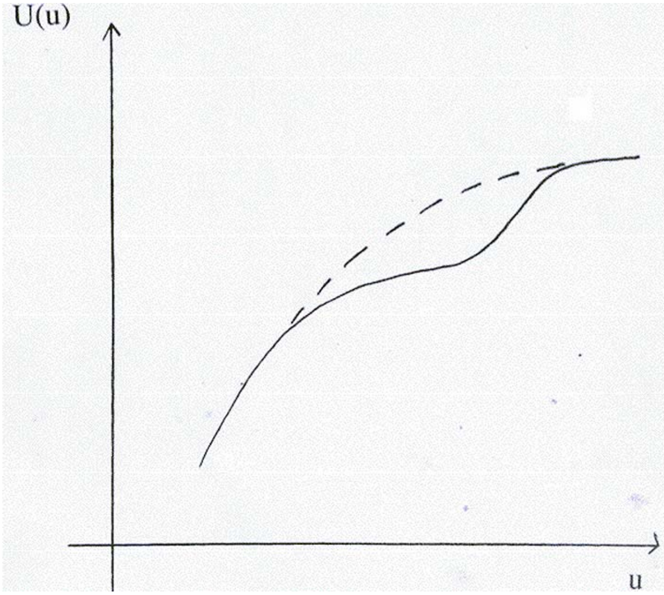


Figure 1

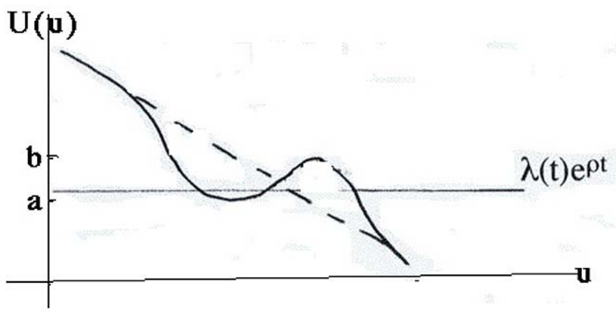


Figure 2

2.6. Jump behaviour with a pseudoconcave objective

An optimal control problem whose objective is pseudoconcave, but not concave, may show jump behaviour in the control, not associated with a boundary of a feasible region. A simple example is proposed in Islam and Craven (2004); it is given here, with numerical results.

Consider the simple example:

$$\text{MAX} \int_0^T U(u(t))dt \text{ subject to:}$$

$$x(0) = x_0 \ , \ \dot{x}(t) = \gamma x(t) - \text{ u(t) } (0 \le t \le T), x(T) = x_T \ .$$

The control $u(.)$ may be unconstrained, or there may be a lower bound

$$(\forall t) u(t) \ge u_{lb}.$$

The horizon T is taken as 10, and the growth factor $\gamma = 1$. The quasiconcave (not concave) utility function $U(.)$ is given by:

$$U(u) = u - 0.5u^2 (0 < u < a),$$

$$U(u) = p + (1-a)(u-a) + 0.5d(u-a)^2 (a < u < b),$$

$$U(u) = q + (1+c)(u-b)\_0.5(u^2 - b^2 \ ) \ (u > b) \ ,$$

with parameters chosen to display the jump effect:

$$a = 0.30, b = 0.35, c = 0.10, p = a - 0.5a^2, d = -1 + c/(b-a),$$

$$q = p + (1-a)(b-a) + 0.5d(b-a)^2.$$

This utility function $U(u)$ is constructed from $U(0) = 0$, and $U'(u) = 1-u, 1-a+d(u-a), 1+b-u$ in the three intervals. The increase in slope makes the function U nonconcave. With these numbers, the nonconcavity of U is only just visible on its graph.

In this example $x(t)$ may represent capital, and $u(t)$ may represent consumption, But, when no lower bound is imposed on $u(t)$, negative values for $u(t)$ may be obtained. The model provides substantial capital growth, and it may be advantageous to borrow money to pay for consumption during the early part of the planning period. To model this better, a cost of borrowing :

$$C(u) = \kappa(u - \delta) \text{ when } u < \delta, C(u) = 0 \text{ when } u \ge \delta,$$

may be subtracted from $U(u)$. This replaces an explicit lower bound on $u(t)$. (Here, $\delta = 0.2$ and $\kappa = 1.3$ were considered.) Also, the dynamic equation for $\dot{x}(t)$ is taken here as linear, to enable an analytic solution to the control problem. A more realistic model would replace $\gamma x(t)$ by some increasing concave function of $x(t)$.

From the theory of Section 2.2, the costate $\lambda(t) = -\sigma e^{-\gamma t}$, with some constant σ depending on the terminal condition for $x(T)$; and then $U'(u(t)) = \sigma e^{-\gamma t}$; hence $u(t)$ increases as t increases. Depending on the initial and terminal conditions and the growth factor γ, the range of variation of the optimal

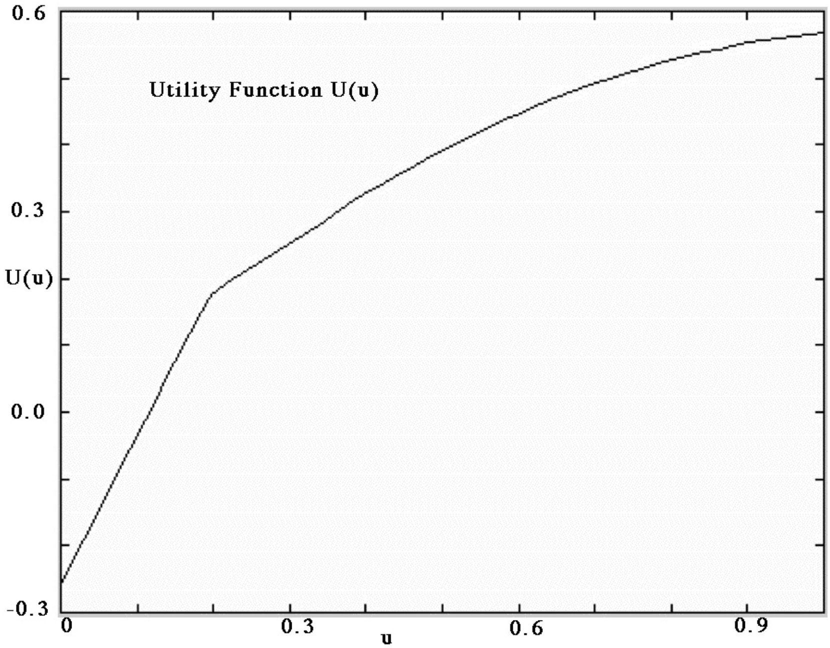


Figure 3 Utility function U(u)

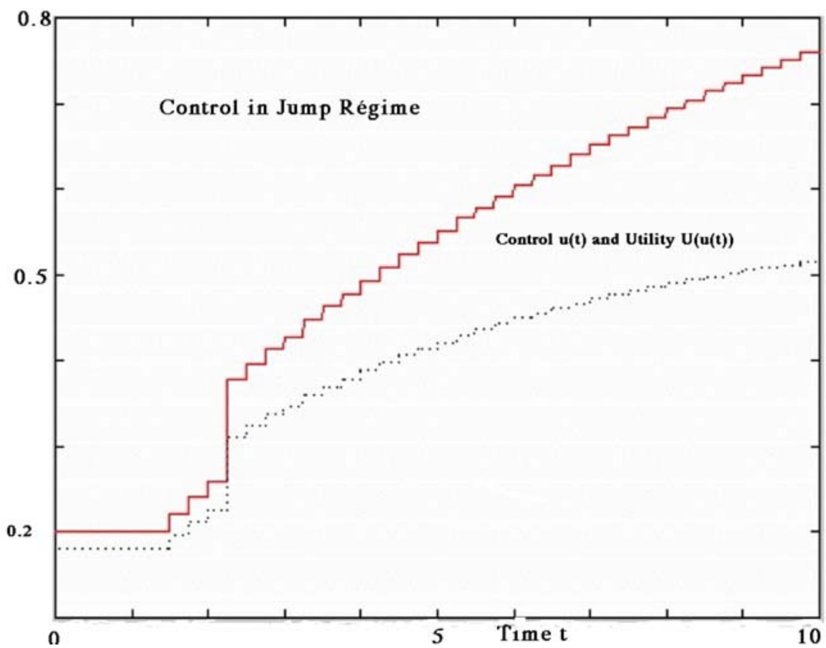


Figure 4 Control in jump regime

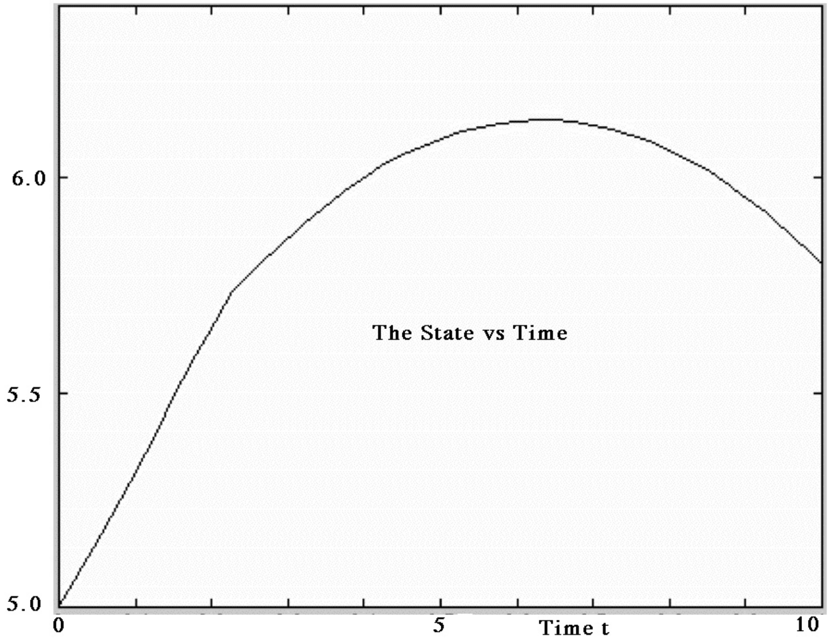


Figure 5 The state

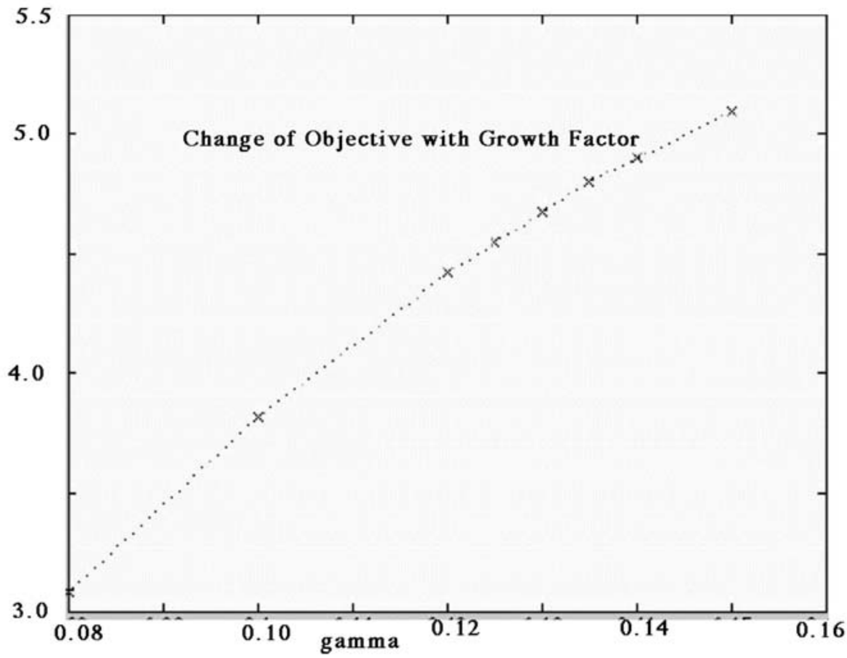


Figure 6 Change of objective with growth factor

$u(t)$ may include the region where there are three possible solutions for $u(t)$. Thus, there are two possible solution regimes, one where this region is not entered, so $u(t)$ varies smoothly, and the second where $u(t)$ jumps from the leftmost part of the $U'(t)$ curve to the rightmost part, at some level of $u(t)$ depending on the parameters mentioned.

Computations with the SCOM optimal control package, described in chapter 3 (also Craven and Islam, 2001) confirm this conclusion. Figure 3 shows the utility function, including the borrowing term. Figure 4 shows the optimal control $u(t)$, and the value of the utility $U(u(t))$, for one set of parameters $(\gamma = 0.10, x0 = 5.0, x_T = 5.80)$. In the computation, $uj(t)$ was approximated by a step-function with 40 subintervals; the theoretical solution would be a smooth curve, except for the jump at $t = 2.3$.

Figure 5 shows the state $x(t)$; note the change in slope at $t = 2, 3$, where $u(t)$ jumps.

Figure 6 shows the change in the optimal objective as the growth factor γ changes, other parameters being the same. Note that the control $u(t)$ changes smoothly when $\gamma \leq 0.13$, whereas $u(t)$ has a jump when $\gamma > 0.13$; the graph in Figure 6 changes slope at this level of γ.

2.7. Generalized duality

While *duality* requires convexity, or some weaker property such as *invex*, there are relaxed versions of maximum for which *quasiduality* holds, giving an analog of ZDG, but not weak duality. A point p is a *quasimax* of $f(x)$, subject to $g(x) \leq 0$, if (Craven, 1977)

$$f(x) - f(p) \leq \mathbf{o}(\|x - p\|) \text{ whenever } g(x) \leq 0.$$

Here a function $q(x) = \mathbf{o}(\|x - p\|)$ when $q(x)/\|x - p\| \to 0$ when$\|x - p\| \to 0$. A function f has a *quasimin* at p if $-f$ has a quasimax at p. If f and g are differentiable functions, and a *constraint qualification* holds, then (Craven, 1997) p is a KKT point exactly when p is a quasimax.

Attach to the problem:

$$\text{QUASIMAX}_x f(x) \text{ subject to } g(x) \leq 0$$

the *quasidual* problem:

$$\text{QUASIMIN}_{u,v} f(u) + vg(u) \text{ subject to } v \geq 0, \; f'(u) + vg'(u) = 0.$$

Then (Craven, 1977, 1995) there hold the properties:

- *ZDG* If x^* is a quasimax of the primal (quasimax) problem, then there is a quasimin point (u^*, v^*) of the quasidual problem for which $f(x^*) = f(u^*) + v^* g f(u^*)$, thus the objective functions are equal.
- *perturbations* if $V(b) = QUASIMAX f(x)$ subject to $g(x) \geq 0$, with the quasimax for $b = 0$ occurring at $x = x^*$, and if (u^*, v^*) are the

optimal quasidual variables for which ZDG hold with $f(x^*)$, then the *quasi-shadow price* vector $V'(0) = v^*$.

There are several quasimax points with related quasimin points, and they correspond in pairs. These *quasi* properties reduce to the usual ones if $-f$ and $-g$ are *invex* with respect to the same *scale function* $\eta(\cdot)$. However, for an equality constraint $k(x) = 0$, the required invex property has $=$ instead of $\geq$.

Simple examples of *quasimax* and *quasidual* are as follows.

Example 1: (Craven, 1977) The *quasidual* of the problem:

$$\text{QUASIMAX}_x \ -x + \frac{1}{2}x^2 \text{ subject to } x \geq 0,$$

is:

$$\text{QUASIMIN}_{u,v} \ -u + \frac{1}{2}u^2 + vu \text{ subject to } -1 + u + v = 0,\ v \geq 0,$$

which reduces to the *quasidual* : $\text{QUASIMIN}_u \ -\frac{1}{2}u^2$ subject to $u \leq 1$. The primal problem has a quasimax at $x = 0$, with objective value 0, and a quasimin at $x = 1$, with objective value $-\frac{1}{2}$. The quasidual has a quasimin at $u = 0$ with objective value 0, and a quasimin at $u = 1$ with objective value $-\frac{1}{2}$.

Example 2: Here, the quadratic objective may be a utility function for social welfare, not necessarily concave, to be maximized, subject to linear constraints. The quasidual vectors are shadow prices. For given vectors c and s, and matrices A and R, suppose that the problem:

$$\text{QUASIMAX}_{z \in \mathbf{R}^n} \ F(z) := -c^T z + (1/2) z^T A z \text{ subject to } z \geq 0,\ Rz \geq s,$$

reaches a quasimax at a point p where the constraint $z \geq 0$ is inactive, and the constraint $Rz \leq s$ is active. Then KKT conditions give $Ap + R^T\lambda = c,\ Rp = s$, with multiplier $\lambda \geq 0$. Setting $z = p + v,\ \ F(z) - F(p) = -\lambda^T Rv - (1/2)v^T Av$ with $Rv \geq 0$. So $-\lambda^T Rv \leq 0$, and $F(z) - F(p)$ may take either sign, depending on the matrix A, so p is generally a quasimax, not a maximum. The associated quasidual problem is:

$$\text{QUASIMIN}_{u,v} \ -c^T u + (1/2)u^T Au - v^T(Tu - s) \text{ subject to}$$

$$c - Au - R^T v = 0,\ v \geq 0.$$

However, the main applicability of *quasimax* is when the primal maximization problem has several local maxima, as is likely to happen when the objective function is far from concave. To each local maximum corresponds a quasimin of the quasidual, with the ZDG property, and quasidual variables, giving shadow prices. Thus the shadow price, for each local maximum, is an optimum point of a quasimin problem.

When there are equality constraints, the *invex* property takes a different form. In general, a vector function F is invex with respect to a convex cone Q at a point p if:

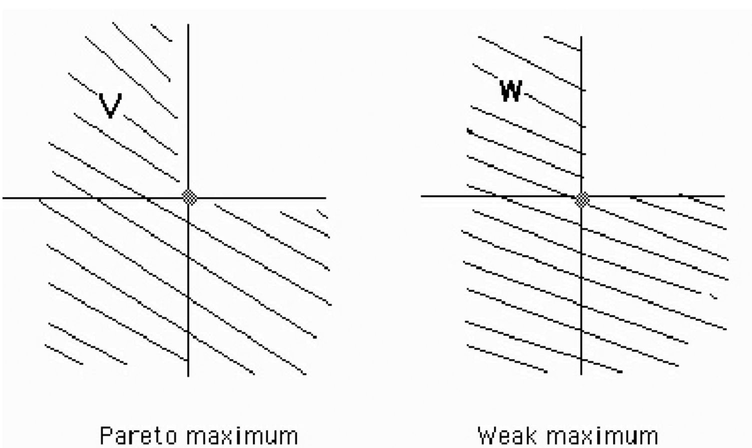
$$(\forall x)\ F(x) - F(p) \in F'(p)\eta(x,p)$$

holds for some *scale function* $\eta(.,.)$. If some component $h(x)$ of $F(x)$ belongs to an equality constraint $h(x) = 0$, then the related part of Q is the zero point $\{0\}$; hence invex requires an equality:

$$(\forall x)\ h(x) - h(p) = h'(p)\eta(x,p).$$

2.8. Multiobjective (Pareto) optimization[5]

When two or more objectives are to be maximized, they usually conflict, so they cannot all be optimal. This is commonly the case with social welfare models. Usually *Pareto* maxima are considered, where moving away from such a point decreases all the objectives. If two objectives F^1 and F^2 should be maximized over a set Γ, define the vector $F(x) := (F^1(x), F^2(x)$. Denote by $\mathbf{W}$ the set of points $w \in \mathbf{R}^2$ for which $w_1 > 0$ and $w > 0$ do not both hold; $\mathbf{W}$ is shown in the right diagram. Then p is a *Pareto maximum* if, for each $x \in \Gamma$, F(x) $-F(p) \in \mathbf{V}$, where $\mathbf{V}$ is the set indicated in the left diagram, namely the interior of $\mathbf{W}$, together with 0. The point p is a *weak maximum* if, for each $x \in \Gamma$, F(x) $-F(p)$ does not lie in $\mathbf{W}$. The weak maximum points include all the Pareto maximum points, and some additional boundary points, excluded from $\mathbf{V}$. For $r > 2$ objectives, $\mathbf{R}^r$ replaces $\mathbf{R}^2$.



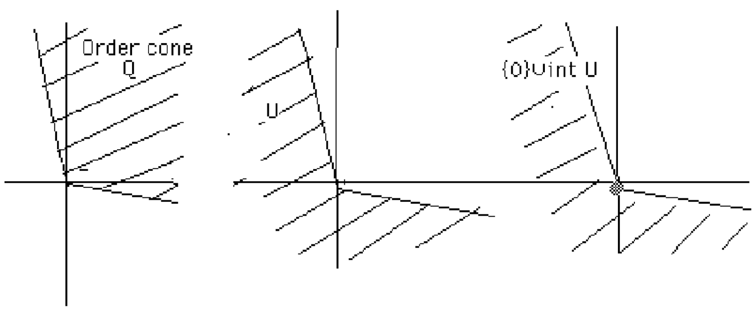
More generally, consider the "*maximization*" of a vector objective function $F(.)$ subject to the constraint $g(.) \geq 0$, relative to an order cone Q in the space

[5] See sections 6.4 and 8.4.

into which F maps. Denote by $\mathbf{U}$ the complement of the interior of Q. A point p is a *weak quasimax* of $F(.)$, subject to $g(.) \geq 0$, if for some function $q(x) = o(||x - p||)$, there holds:

$$F(x) - F(p) - q(x) \in \mathbf{U} \text{ whenever } g(x) \geq 0.$$

(Note that *weak maximum* is the special case $q(.) = 0$.) Replacing $\mathbf{U}$ by the union of 0 with the interior of $\mathbf{U}$ gives the slightly more restrictive *Pareto maximum;* a *weak maximum* may include a few additional boundary points. A *weak quasimin* of a vector function H is a weak quasimax of -H . If F has r components, then usually Q is taken as the orthant $\mathbf{R}^r_+$. The following diagram (see Craven, 1981) illustrates the definition, for the case when F has two components, so that the order cone Q is a sector, but not always the non-negative orthant $\mathbf{R}^2_+$. Such an ordering is appropriate when the consumers are not isolated, as is discussed in section 8.2.



From Craven (1995), a (maximizing) *weak Karush-Kuhn-Tucker point (WKKT)* p satisfies:

$$\tau F'(p) + \lambda g'(p) = 0, \lambda g(p) = 0, \lambda \geq 0\ \tau \geq 0, \tau e = 1.$$

Here τ is a row vector, e is a column vector of ones. The statement $\tau \geq 0$ means, more precisely, that τ lies in the dual cone Q^* of Q, defined by $q^*(q) \geq 0$ whenever $q^* \in Q^*$ and $q \in Q$, If $Q = \mathbf{R}^r_+$ then each component of τ is non-negative. There are many WKKT points, corresponding to different values of τ.

Analogously to section 2.7 (and see Craven, 1990), the *weak quasidual* to the (primal) problem:

$$\text{WEAK QUASIMAX}_x F(x) \text{ subject to } g(x) \geq 0$$

has the form:

$$\text{WEAK QUASIMIN}_{u,V} F(u) + Vg(u) \text{ subject to}$$

$$V(\mathbf{R}^m_+) \subset Q, (F'(u) + Vg'(u))(X) \subset \mathbf{U},$$

where the space $X = \mathbf{R}^n$ is the domain of F and g, and V is a $r \times m$ matrix variable. The *linearized problem* about the point p:

$$\text{QUASIMAX } F'(p)(x-p) \text{ subject to } g(p) + g'(p)(x-p) \geq 0$$

If a constraint qualification holds, and without any convex or invex assumption, there hold (Craven, 1989, 1990):

- p is a WKKT point exactly when p is a weak quasimin.
- *ZDG* If x^* is a weak quasimax point of the primal, with multipliers τ, λ, then there there is a weak quasimax point $(u, V) = (x^*, V^*)$ of the weak quasidual problem, with $\tau V^* = \lambda$, and equal optimal objectives:

$$F(x^*) = F(u^*) + V^* g(u^*).$$

Proof Let x and u, V be feasible for the respective problems. Then (using quasidual constraints):

$$[F(u) + Vg(u)] - [F(x^*) + V^* g(x^*)) \in Q + [F(u) + V^* g(u)]$$

$$-[F(x^*) + V^* g(x^*) + (V - V^*)(g(u) - g(x^*)\}$$

$$= Q + (F + Vg)'(x^*)(u - x^*) + o(\|u - x^*\|) + (o(\|u - x^*\|) + o(\|V - V^*\|)),$$

$$\subset \mathbf{U} + o(\|u - x^*\| + \|V - V^*\|).$$

- *Linearization* The linearized problem about p reaches a weak maximum at p.
- *Shadow prices* If x^* is a weak quasimax of the primal, then x^* is also a quasimax of $\tau^* F(x)$; so the shadow prices for $\tau^* F$ are optimal for:

$$\text{QUASIMAX } F'(p)(x-p) \text{ subject to } g(p) + g'(p)(x-p) \geq 0$$

- *Multilinear problem* (Bolinteanu and Craven, 1992) If F is linear, and g is linear (plus a constant), then p is stable to small perturbations, and there is a shadow price for F for perturbations that do not change the list of active constraints.

Example 3 - Vector quasimax and quasidual

$$\text{WEAK QUASIMAX}_{z \in \mathbf{R}^n} \{F_i(z)\}_{i=1}^r := \{-c_i^T z + (1/2) z^T A_i iz\}_{i=1}^r,$$

$$\text{subject to } Rz \geq s,$$

for given vectors c_i, d, s and matrices A_i, R; here $\{F_i(z)\}_{i=1}^r$ specifies a vector objective by its components. For each multiplier $\tau \geq 0$, with $e^T \tau = 1$, where e is a vector of ones, there corresponds a quasimax $z = z(\tau)$, with:

$$A^T z(\tau) + R_0^T \lambda(\tau) = c, \;\; R_0 z(\tau) = s_0,$$

in which $R_o z \geq s_0$ describes the constraints active at $z(\tau)$, $c = \sum \tau_i c_i$, $A = \sum \tau_i A_i$, As in example 2, this quasimax becomes a maximum if the matrix A is restricted, e.g. to be negative definite in feasible directions. This problem may describe e.g. conflicting objectives of output and envoronmental quality for sustainable development, and requiring nonlinear functions to describe them. The vector quasidual is then:

$$\text{WEAK QUASIMAX}_{u,V} \; \{-c_i + u^T A_i\}_{i=1}^r + VR)(\mathbf{R}^n) \subset \mathbf{U}, \; V \geq 0,$$

where $\mathbf{U} = \mathbf{R}^r \backslash$ int $\mathbf{R}_+^r$; each component of V is ≥ 0.

While a *quasimax* is not generally a maximum, the *quasi* properties reduce to the standard properties if an additional assumption, such as *convex* or *invex*, is made. Consider the problem of weak maximization of $F(x) <$ *with* respect to the order cone Q, subject to constraints $g(x) \geq 0$ and $k(x) = 0$. In weak KKT, a term $\sigma k'(p)$ is added to $\lambda g'(p)$, where the multiplier σ is not restricted in sign. The *invex* property stated in section 3.2 (see Hanson, 1980; Craven, 1981; Craven, 1995) now takes the form:

$$-F(x) + F(p) + F'(p)\eta(x,p) \in Q,$$

$$g(x) - g(p) \leq g'(p)\eta(x,p),$$
$$k(x) - k(p) = k'(p)\eta(x,p),$$

for the same *scale function* $\eta(x,p)$ in each case. (The inequalities derive from a *cone-invex* property $H(x) - H(p) - H'(p)\eta(x,p) \in Q \times \mathbf{R}_+^m \times \{0\}$ for a vector function $H = -(F, g, k)$ (see Craven 1995).

In particular, if all components of F and g are concave, and all components of k are affine (constant plus linear), then invexity holds with η (x, p) $= x - p$, and the necessary conditions *weak KKT*, or equivalently *weak quasimax,* become also sufficient for a maximum. But these assumptions are often not fulfilled in applications. Some conditions when invexity holds are given in Craven (1995, 2000). A transformation $x = \varphi$ (y) of the domain, where φ is invertible, with both φ and φ^{-1} differentiable, preserves the *invex* property, although the scale function is changed. The name *invex* derives (Craven, 1981) from *invariant convex* ; thus if a function $\mathbf{H}$ is convex, then the transformed function $\mathbf{H} \circ \varphi$ is *invex* (though not every invex function is of this form). However, for an equality constraint $k(\ \mathbf{x}\ ) = 0, k(.)$ is invex at p if $k(x) - k(p) = k'(p)\eta$ (x, p), For KKT to be sufficient for an optimum, this invex property is only required for those x where $k(x) = 0$. So the requirement reduces to the *reduced invex (rinvex)* property: $k(x) = 0 = k'(p)\eta(x,p)$.

A multiobjective analog of *saddlepoint* was given in Craven (1990), and the results are summarized here. The point p is a *weak saddlepoint* (WSP) (see Craven, 1990) if:

$$\tau^T F(p) + v^T g(p) \geq \tau^T F(p) + \lambda^T g(p) \geq \tau^T F(x) + \lambda^T g(x)$$

holds for all $\mathbf{x}$, and all $v \geq 0$. The multiplier τ (where $0 \neq \tau \geq 0$) depends on p. The point p is a *weak quasisaddlepoint* if (given τ):

$$(\forall x, \forall \mu \geq 0)\tau F(p) + \mu g(p) \geq \tau F(p) + \lambda g(p) \geq \tau F(x) + \lambda g(x) + \mathbf{o}(\|x - p\|).$$

The following relations hold (Craven, 1995) for the vector function F on E and $p \in E$:

p is a Pareto maximum $\Rightarrow p$ is a weak maximum

$\Rightarrow p$ is a weak quasimax

$\Leftrightarrow p$ is a weak KKT point (assuming regularity of the constraint)

$\Leftrightarrow p$ is a weak quasisaddlepoint .

Weak maximum (and *weak quasimax*) points may include some boundary points which lack a desired stability property. The statement that $0 = \tau = (\tau_1, \ldots, \tau_r\ ) \geq 0$ means that all components $\tau_i \geq 0$, and they are not all zero. If $\tau > 0$, namely if all $\tau_i > 0$, then (Craven, 1990): weak KKT & $\tau > 0 \Leftrightarrow$ *locally proper weak quasimax* , where *locally proper* means that ratios of deviations of the different $\mathbf{F}_i$ from $\mathbf{F}_i p)$ are bounded (inequalities of Geoffrion, 1968) when x is near p. These ratios represent *tradeoffs* between the objective components. They are approximated by the corresponding ratios of the τ_i. Thus a *proper* quasimax occurs exactly when all the multipliers τ_i for the objectives are strictly positive. This property does not require the Lagrange multipliers to be stated. As motivation, note that the *stationary points* of a real function f are the point p where the gradient $f'(p) = 0$; and maximum points are sought among the stationary points, If there are constraints, then *constrained maximum points* are among the KKT points instead of stationary points, A quasimax point p is a maximum point of the linear program, obtained by linearizing the given problem about the point p.

2.9. Multiobjective optimal control

For an optimal control problem with a single objective, the KKT conditions are essentially equivalent to an adjoint differential equation for a costate function, together with Pontryagin's principle. This extends to multiobjective control problems.

Consider an optimal control problem, to find Pareto maximum points of a vector objective $F(x, u)$, subject to a dynamic equation that determines the state x in terms of the control u, and other constraints. By substituting for x in terms of u, the problem takes the form:

$$\text{PMAX } \mathbf{J}(u) \text{ subject to } u \in \mathbf{U},$$

where PMAX denotes Pareto maximum, and the feasible set $\mathbf{U}$ is a closed bounded subset of some vector space $\mathbf{V}$. Suppose that $u \in \mathbf{U} \Leftrightarrow \mathbf{G}(u) \leq 0$, for some function $\mathbf{G}$. If a Pareto maximum point $\hat{u}$ is reached, then (under restrictions usually satisfied), multipliers τ and ρ exist such that Karush-Kuhn-Tucker (KKT) conditions hold:

$$\tau\mathbf{J}'(\hat{u}) + \rho\mathbf{G}'(\hat{u}) = 0, 0 \neq \tau \geq\ 0,\ \rho \geq\ 0,\ \rho\mathbf{G}(\hat{u}) = 0.$$

If $-\mathbf{J}$ and $\mathbf{G}$ satisfy some relaxed version of convexity, such as *invexity* (see Craven, 1995), then the KKT conditions, in turn, imply a Pareto maximum, and also that $\tau\mathbf{J}(u)$ is maximized, subject to $\mathbf{G}(u) \leq 0$, at $\hat{u}$. Then each τ corresponds to one or more Pareto maximum points, and each Pareto maximum point corresponds to one or more multipliers τ. If minimum points of $\tau\mathbf{J}(.)$, subject to $\mathbf{G}(u) \leq 0$, exist for each $\tau \geq 0$, then the existence of Pareto maximum points is established.

For an optimal control problem in continuous time, the vector space $\mathbf{V}$ is infinite dimensional, so compactness properties of $\mathbf{U}$ are not generally available, so a Pareto maximum of a continuous vector function is not always reached. For each given τ, assume that $J(u) := \tau\mathbf{J}(u)$ has a finite lower bound over $\mathbf{U}$. If $J(u)$ has the form $\int_0^T f(x(t), u(t), t)dt$, with a *coercivity* assumption on the integrand f, that:

$$(\forall x, u, t) f(x, u, t) \geq g(u)\ ,\ \text{where } g(u)/|u| \to \infty \text{ as } |u| \to \infty,$$

together with convexity of a set: $\cup_{u\in U}\{$ (z, y) : $z = m(x, u, t)y \geq f(x, u, t)$, where U is the closed set of values allowed to $u(t)$, then a maximum for $J(u)$ exists (see Fleming and Rishel, 1975, Theorem 4.1). But these assumptions are often not satisfied for an economic growth model.

Suppose, however, that $u = p$ is a solution of the KKT necessary conditions for a maximum of $J(u)$ (or equivalently of the Pontryagin conditions – see section 2.2). Additional assumptions are needed, to ensure that p is a maximum. Suppose that $S_1 \subset S_2 \subset \ldots \subset S_n \subset \ldots$ is an increasing sequence of finite-dimensional subspaces of $\mathbf{V}$, and that p is a limit point of their union. Suppose also that each $S_n \cap \mathbf{U}$ is bounded (in a finite dimension); then $J(u)$ reaches a maximum, say at $u = p_n$, over $S_n \cap \mathbf{U}$. Then $J(p_n) \to J(p)$, and p is a maximum point for J. (See Craven (1999b) for an example for a control model where bang-bang control is optimal.)

Otherwise, if $u = p$ satisfies the Pontryagin conditions, and if the functions of the problem satisfy *invex* conditions (see Craven 1995), then p is a maximum. In particular, it suffices if the integrand f of the objective is concave, and the dynamic differential equation for the state is linear. Less restrictively, a function φ is *invex* at p if :

$$\varphi(u) - \varphi(p) \geq \varphi'(p)\eta(u, p),$$

where the *scale function* η must be the same for the several functions (-f and constraint functions) describing the problem. But this is often not the case for a given dynamic equation. However, the control problem has a *local* maximum at p if $-f$ is invex, under lesser restrictions on the differential equation (see section 8.6.)

The maximum point p for $J(u)$ is unique if the Pontryagin conditions have a unique solution. This may happen, in particular (see Islam and Craven, 2001b) when there are no control constraints, so that the Pontryagin maximum condition reduces to a gradient equation, $f_u + \lambda m_u = 0$. If this equation defines u as a unique function of λ, and some Lipschitz conditions hold for f, m, m_x and m_u , then the differential equations for $\dot{x}$ and $\dot{\lambda}$ (substituting for u) have unique solutions.

For a multiobjective problem, the maximum point cannot be unique, but the set of Pareto maximum points is unique, if one of the above criteria applies.

2.10. Multiobjective Pontryagin conditions[6]

Consider an optimal control problem, to find a Pareto (or weak) maximum for two objective functions over a time interval [0, T] :

$$\text{PMAX } \{\int_0^T f^1(x(t), u(t), t)dt + \Phi^1(x(T)),$$

$$\int_0^T f^2(x(t), u(t), t)dt + \Phi^2(x(T))\}$$

subject to:

$$x(0) = x_0, \dot{x}(t) = m(x(t), u(t), t) \ (0 \ \leq t \leq T),$$

$$a(t) \leq u(t) \leq b(t) \ (0 \ \leq t \leq T).$$

Here the state function $x(.)$ and control function $u(.)$ can be scalar or vector valued; PMAX denotes Pareto maximum; there may be endpoint contributions Φ_1, Φ_2. Note that a Pareto minimum of $\{F^1, F^2\}$ is a Pareto maximum of $\{-F^1, -F^2\}$. There can be more than two objectives.

The relevance of such models to economic growth, decentralization, social choice, and sustainable growth is discussed in chapters 5, 6 and 8. In particular, components of $u(.)$ may represent consumption, and components of $x(.)$ may represent capital stock, with a dynamic differential equation describing economic growth. Questions of existence and uniqueness are discussed in sections 2.3 and 2.5.

The above optimal control model can be described by a *Hamiltonian* :

$$h(x(t), u(t), t; \tau, \lambda(t)) :=$$

[6] See section 8.4.

$$\sum_i \tau_i[f^i(x(t),u(t),t)+\delta(t-T)\Phi^i(x(T))]+\lambda(t)m(x(t),u(t),t)$$

Here $\lambda(.)$ is the *costate function* ; τ_i are nonnegative multipliers; and the end-point terms $\Phi^i(x(T))$ have been replaced by terms $\delta(t-T)\Phi^i(x(T))$ added to the integrands; $\delta(.)$ is Dirac's delta-function.

The analysis in Craven (1999a) uses a *Lagrangian* $L(.)=h(.)+\dot{\lambda}(t)x(t)$.

The optimum is then described by the necessary Karush-Kuhn-Tucker (KKT) conditions. From this is deduced, under some regularity assumptions, a *(weak) adjoint differential equation*:

$$\dot{\lambda}(t)=\tau^T f_x(x(t),u(t),t)+\lambda(t)m_x(x(t),u(t),t),\lambda(T)=\tau^T\Phi'(x(T)),$$

(where f_x denotes $(\partial/\partial x)f$, etc.), together with the Pontryagin principle:

$$h(x(t),.,t;\tau,\lambda(t))\rightarrow\text{ MAX over }[a(t),b(t)]\text{ at the optimal }u(t).$$

There are many Pareto (or weak) optimal points, corresponding to various values of τ, where $0\neq\tau\geq 0$. If the constraints on the controls are inactive, then MAX h implies:

$$h_u(x(t),u(t),t;\tau,\lambda(t))=0$$

at the optimal $x(t),u(t),\lambda(t)$. (Note that this does not imply MAX h.)

The costate function $\lambda(t)$ represents a *shadow price* for the differential equation. Thus, if the dynamic differential equation is perturbed to:

$$x(0)=x_0,\dot{x}(t)=m(x(t),u(t),t)-\beta(t),$$

where $\beta(.)$ is a continuous function of small norm, then the change in the optimal value of $\tau f(.)$ is approximated by $\int_0^T\lambda(t)\beta(t)dt$.

An alternative description uses a *vector Hamiltonian* :

$$\mathbf{H}(x(t),u(t),t;\Lambda(t)):=$$

$$\mathbf{f}(x(t),u(t),t)+\delta(t-T)\Phi(x(T))+\Lambda(t)m(x(t),u(t),t),$$

in which $\mathbf{f}$ is the vector of f^i,Φ is the vector of φ^i , and $\Lambda(t)$ is a matrix valued function instead of the vector function $\lambda(.)$. The *vector adjoint equation* is then:

$$\dot{\Lambda}(t)=f_x(x(t),u(t),t)+\Lambda(t)m_x(x(t),u(t),t),\Lambda(T)=\Phi'(x(T)).$$

Corresponding to Pontryagin's principle, there is a Pareto (or weak) optimum of the vector Hamiltonian, thus:

$$\mathbf{H}(x(t),.,t;\Lambda(t))\rightarrow\text{ PMAX over }[a(t),b(t)]\text{ at the optimal }u(t).$$

The *maximum* in Pontryagin's principle becomes here a *Pareto maximum.* If the constraints on the controls are inactive, then optimizing the vector Hamiltonion implies (for two objectives) that:

$$(\exists \tau \geq 0, \tau \neq 0)\tau \mathbf{H}_u(x(t), u(t), t; \Lambda(t)) = 0,$$

at the optimal $x(t), u(t), \Lambda(t)$. (Note that this does not imply PMAX.)

If multipliers τ and λ are known, then $\Lambda(t)$ may be constructed from $\tau^T \Lambda(t) = \lambda(t)$.

http://www.springer.com/978-0-387-24280-4

Optimization in Economics and Finance
Some Advances in Non-Linear, Dynamic, Multi-Criteria
and Stochastic Models
Craven, B.D.; Islam, S.M.N.
2005, XI, 161 p.,
ISBN: 978-0-387-24280-4