ALGORITHMS FOR NONLINEAR PROGRAMMING AND MULTIPLE OBJECTIVE DECISIONS

(J. Wiley, 1998; ISBN 0 471 97850 7)

Berç Rustem
Department of Computing
Imperial College of Science, Technology & Medicine
180 Queen's Gate, London SW7 2BZ, UK

PREFACE

This book is a study of algorithms for decision making with multiple objectives. It is addressed to researchers in computational methods for decision making and optimal design: computer scientists interested in quantitative decision support, in particular numerical optimization; engineers; mathematicians; and those working in management science; operations research; economics and finance. Although most of the mathematics required is reviewed in a series of comments and notes, located at the end of relevant chapters, the reader is expected to have developed some insight in decision making and associated concepts.

The starting point is the nonlinear optimal decision problem for dynamic systems with multiple objectives under uncertainty. An optimal decision needs to take account of possible future uncertainties. As the process unfolds, and the uncertainty becomes known, the decision is revised and new future uncertainties are considered. This approach to optimal decisions is formulated as a static nonlinear problem and the question of multiple objective decision making within this framework is considered using quadratic programming, nonlinear programming, nonlinear constrained min-max, mean-variance optimization and noncooperative Nash games. Regarding uncertainty in multi-period decision problems, the treatment of scenario optimization is omitted from this book. This approach utilises the probability of the scenarios at each period to evaluate the expected value of the objective and ensure the satisfaction of the constraints arising from each scenario. It is essentially the application of optimization algorithms to very large scale problems. The size of the problem is in particular due to the scenarios and adds a further layer of complexity, often resolved by efficient problem formulation, decomposition and parallel computers. Other areas not covered in this book are network optimization, combinatorial optimization and integer programming. The algorithms and tools for these seem to be beyond the main focus of the book.

We consider the static optimization problem with a single criterion in chapter 1 and study the optimality conditions under equality and inequality constraints. We also describe in chapter 1, a simple and approximate algorithm for solving the nonlinear dynamic policy optimization problem. If an approximate search strategy is required, when the multiple objective decision is being formulated, the last two sections of chapter 1 provide an algorithm that has been tried and tested in numerous macroeconomic decision making and engineering process control problems. Nevertheless, it must be pointed out that the rest of the book is devoted to the discussion of methods for which accuracy is of primary importance.

Chapter 2 is devoted to the solution of the quadratic programming problem, encountered in chapters 3-5, for the specification of multiple objective problems and, as a subproblem, in chapters 6, 7, 8 and 12. Three different algorithms are considered in this chapter.

The basic view of multiple criteria is that the decision making process is a cognitive one. It is in the course of this process that the decision maker gains an increasingly concrete knowledge of what can be done and determines the trade-offs and targets. chapters 3-5 describe iterative methods, involving interactions with the decision maker, for the specification of the relative weights and targets in multiple objective problems. In chapter 5, the convergence properties of these algorithms are considered.

Chapters 6-8 cover nonlinear programming algorithms required for the solution of the optimization of a single objective. Convex optimization is discussed in chapter 6 and an efficient version of the Goldstein-Levitin-Polyak algorithm is studied in detail with convergence rate results. The general nonlinear programming problem is considered in chapter 7 with a detailed study of sequential quadratic programming algorithms. For example, considerations such as convexifying the problem in order to enlarge the region of convergence of the algorithm and augmented Lagrangians are introduced. Techniques for augmenting the Lagrangian are discussed along with stepsize strategies that measure the progress of the algorithm at every iteration. The rates of convergence of these algorithms are dependent on the nature of the approximate Hessian used. These are discussed in chapter 8, with results concerning the rates for the variable-Lagrange multiplier pair.

A competition model in the presence of multiple decision makers is considered in chapter 9. This is the case when each objective corresponds to an agent, or player, whose actions affect the system and thereby the objectives of other players. The aim is the computation of Nash equilibria in games. The algorithms considered are: an asynchronous relaxation of the best replay algorithm, as well as variants of the Newton algorithm for solving the equilibrium condition.

The mean-versus-variance multiple objective problem is discussed in chapters 10-11. This arises in decision making under uncertainty where we consider the simultaneous optimization of the expected value of the objective and its variance, which represents the associated risk. These are usually conflicting objectives. An example is the classical investment portfolio problem of maximizing expected return and minimizing expected risk. This is studied in chapters 10-11. In the former, an extension of the portfolio problem is discussed, combining the risk of investments and exchange rates. In the latter, a study of the nonlinear case is given for dynamic decision problems with feedbacks.

The final approach to the multiple objective problem, discussed in chapter 12, is the min-max formulation. The optimization of the worst-case objective requires an algorithm to solve the nonlinearly constrained min-max problem. The algorithm and its convergence rate properties are considered in detail.

The ideas presented in this book are based on experience in designing solutions to optimal decision problems in economics, finance and engineering. These were developed over a period during which I was privileged with the opportunity to discuss, debate and argue related questions with Robin Becker, Jeremy Bray, Kumaraswamy Velupillai. Without their input, parts of the book would have been considerably weaker. On nonlinear programming, I am indebted to Laurence Dixon and David Mayne for numerous informative discussions and to Ioannis Akrotirianakis for proof reading most of the related chapters. On various aspects of uncertainty, I am grateful for the comments and advice of Gregory Chow, David Kendrick and Martin Zarrop. Of course, none of the above bear responsibility for any remaining errors or misrepresentations.

The book was finished during a sabbatical year and I am grateful to Imperial College for the sabbatical programme and to my colleagues in the Department of Computing for giving me this opportunity to finalise the project.

ALGORITHMS FOR NONLINEAR PROGRAMMING & MULTIPLE OBJECTIVE DECISIONS

Berç Rustem
Department of Computing
Imperial College of Science, Technology & Medicine
180 Queen's Gate, London SW7 2BZ, UK

CONTENTS

NOTATION

Sections within a chapter are referred by their consecutive numbers, sections in other chapters are preceded by the chapter number. For example, in chapter 1, the third section is referred to as section 3 whereas within chapter 2, the same section would be referred to as section 1.3. The numbering of equations, theorems etc also follow the same rule. Within a chapter, an equation is referred to by the section number in which it occurs, followed by the the equation or theorem number in the section. Outside the chapter, this is preceded by the chapter number. Although sometimes subsections are used to direct the discussion, these are not used in the referencing system.

$t = 1, 2, ..., \mathcal{T}$:	Index of discrete-time periods, starting at period 1 with final time period $\mathcal{T}$.
$u_t \in \mathbb{R}^u$:	Vector of controls, or decision variables at t^{th} time period.
$y_t \in \mathbb{R}^y$:	Vector of output, or endogenous values, determined by the system, at t^{th} time period.
$\epsilon_t \in \mathbb{R}^\epsilon$:	Vector of uncertainties, or random variables, at t^{th} time period.
$\mathcal{E}(.)$:	Expected value of (.).
var (.):	Variance of (.).
$U \equiv [u_1^T, ..., u_t^T, ..., u_{\mathcal{T}}^T]^T \in \mathbb{R}^{u \times \mathcal{T}}$:	Vector of controls of all time periods.
$Y \equiv [y_1^T, ..., y_t^T, ..., y_{\mathcal{T}}^T]^T \in \mathbb{R}^{y \times \mathcal{T}}$:	Vector of endogenous variables of all time periods.
U^d, Y^d:	Desired, or bliss values of U, Y.
$\epsilon \equiv [\epsilon_1^T, ..., \epsilon_t^T, ..., \epsilon_{\mathcal{T}}^T]^T \in \mathbb{R}^{\epsilon \times \mathcal{T}}$:	Vector of random variables of all time periods.
$\mathbb{R}^n$:	n dimensional real vector space.
$x \in \mathbb{R}^n$:	Vector of optimization variables. Sometimes we denote $x = [Y^T \vdots U^T]^T$.
x^d:	Desired, or bliss, value of x.
x_p:	Preferred value of x (chapters 3-5).
x_c:	Current optimal value of x (chapters 3-5).
x_n:	New optimal value of x (chapters 3-5).
$\bar{x}$:	Unconstrained Newton step at x_k (chapter 6).
x_k^p:	Projection of $\bar{x}$ (chapter 6).
d_k:	Direction of search at x_k.
τ_k:	Stepsize along direction of search d_k.
c_k:	Penalty parameter value at x_k.
η_k:	Barrier parameter at iteration k.
L:	Binary input length of quadratic programming problem.
$\mathbb{R}^{n \times m}$:	Set of real matrices of dimensions $n \times m$.
$\mathbb{R}_+^m \equiv \left\{ \eta \in \mathbb{R}^m \mid \eta \geq 0 \right\}$:	Nonnegative orthant.
$\mathit{1} \in \mathbb{R}^n$:	$[1, 1, ..., 1]^T$
$< x, y >$:	Inner product $x^T y$ for $x, y \in \mathbb{R}^n$.
$\mathbb{E}_+^m \equiv \left\{ \alpha \in \mathbb{R}^m \mid \alpha \geq 0 ; < \mathit{1}, \alpha > = 1 \right\}$	
f(x):	Scalar objective function of $x \in \mathbb{R}^n$.
$f \in \mathbb{C}^1(\mathbb{R}^n)$, or $\in \mathbb{C}^1$:	Function f has continuous first partial derivatives with respect to $x \in \mathbb{R}^n$.
$f \in \mathbb{C}^2$:	Function f has continuous second partial derivatives with respect to x.
$\nabla_x f(x)$, or $\nabla f(x)$:	Gradient of f with respect to x. This is a column vector with $(j)^{th}$ element $\frac{\partial f(x)}{\partial x^j}$:

$$\nabla f(x) \equiv \begin{bmatrix} \frac{\partial f(x)}{\partial x^1} \\ \vdots \\ \frac{\partial f(x)}{\partial x^j} \\ \vdots \\ \frac{\partial f(x)}{\partial x^n} \end{bmatrix}$$

f_k and ∇f_k : $f(x_k)$ and $\nabla f(x_k)$

$q(x)$, or $q_k(x)$: quadratic objective function or quadratic approximation to the objective function at x_k.

$\mathcal{Y}(U)$: computational mapping between Y and U arising from model of the system $g(Y, U) = 0$ and the model solution algorithm (chapters 1, 11).

$\mathcal{F}(x)$, or $\mathcal{F}(U)$: $f(\mathcal{Y}(U), U)$, i.e. the objective function $f(Y, U)$ reduced using the mapping $Y = \mathcal{Y}(U)$.

$\mathcal{Q} \equiv \nabla^2 f(x)$: Hessian, or second derivative matrix, of f with respect to x. $(j, l)^{th}$ element of this matrix is given by $\frac{\partial^2 f(x)}{\partial x^j \, \partial x^l}$:

$$\nabla^2 f(x) \equiv \begin{bmatrix} \frac{\partial^2 f(x)}{\partial (x^1)^2} & \cdots & \frac{\partial^2 f(x)}{\partial x^j \, \partial x^1} & \cdots & \frac{\partial^2 f(x)}{\partial x^n \, \partial x^1} \\ \vdots & \ddots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x^1 \, \partial x^j} & \cdots & \frac{\partial^2 f(x)}{\partial (x^j)^2} & \cdots & \frac{\partial^2 f(x)}{\partial x^n \, \partial x^j} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x^1 \, x^n} & \cdots & \frac{\partial^2 f(x)}{\partial x^j \, \partial x^n} & \cdots & \frac{\partial^2 f(x)}{\partial (x^n)^2} \end{bmatrix}$$

$\mathcal{Q}_u$, $\mathcal{Q}_y$: Weighting matrices associated with $U - U^d$, $Y - Y^d$

$g(x) \in \mathbb{R}^e$: Vector valued function of equality constraints $\left(\text{i.e. } \left\{ x \in \mathbb{R}^n \mid g(x) = 0 \right\} \right)$.

g_k: $g(x_k)$

∇g_k: $\nabla g(x_k) \equiv [\nabla g^1(x_k) \vdots \nabla g^2(x_k) \vdots \ldots \vdots \nabla g^e(x_k)] \in \mathbb{R}^{n \times e}$

$h(x) \in \mathbb{R}^i$: Vector valued function of inequality constraints $\left(\text{i.e. } \left\{ x \in \mathbb{R}^n \mid h(x) \leq 0 \right\} \right)$.

h_k: $h(x_k)$

∇h_k: $\nabla h(x_k) \equiv [\nabla h^1(x_k) \vdots \nabla h^2(x_k) \vdots \ldots \vdots \nabla h^i(x_k)] \in \mathbb{R}^{n \times i}$

$\mathcal{I}(x_k) \equiv \left\{ i \mid h^i(x_k) = 0 \right\}$: Set of active inequality constraints at x_k.

$\mathcal{R}$: Set of feasible points, usually given by $\mathcal{R} = \left\{ x \in \mathbb{R}^n \mid g(x) = 0 ; \; h(x) \leq 0 \right\}$.
In chapter 6, where we constrain $\mathcal{R}$ to be convex, we have: $\mathcal{R} = \left\{ x \in \mathbb{R}^n \mid h(x) \leq 0 \right\}$.
In chapter 3, we take this set to be defined by linear equalities only:
$\mathcal{R} = \left\{ x \in \mathbb{R}^n \mid \mathcal{G}^T x = g \right\}$.

Ω: Set of policies acceptable to the decision maker (chapters 3-5)

$L(x, \lambda, \mu)$: Lagrangian function for the constrained optimization problem.

$L^a(x, \lambda, \mu, c, \alpha)$:	Augmented Lagrangian function for the constrained optimization problem.
$\alpha \in \mathbb{R}^i$:	Vector of offsets for inequality constraints, used in chapters 7 and 8.
$\lambda \in \mathbb{R}^e$:	Multiplier vector for the equality constraints.
$\mu \in \mathbb{R}^i$:	Multiplier vector for the inequality constraints.
trace (A):	$\sum_{i=1}^{n} a_{ii}$; $A \in \mathbb{R}^{n \times n}$
diag (x):	$\text{diag}\,[x^1, x^2, ..., x^n] \equiv \begin{bmatrix} x^1 & 0 & \cdots & 0 \\ 0 & x^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x^n \end{bmatrix}$; $x \in \mathbb{R}^n$
$\lceil x \rceil$:	smallest integer ι such that $\iota \geq x$.
□:	End of a proof, an example, or a particular train of thought.
policy maker:	decision maker