1. WHAT IS OPTIMIZATION?

Optimization problem: Maximizing or minimizing some function relative to some set, often representing a range of choices available in a certain situation. The function allows comparison of the different choices for determining which might be "best."

Common applications: Minimal cost, maximal profit, minimal error, optimal design, optimal management, variational principles.

Goals of the subject: The understanding of

Modeling issues—

What to look for in setting up an optimization problem?

What features are advantageous or disadvantageous?

What devices/tricks of formulation are available?

How can problems usefully be categorized?

Analysis of solutions—

What is meant by a "solution?"

When do solutions exist, and when are they unique?

How can solutions be recognized and characterized?

What happens to solutions under perturbations?

Numerical methods—

How can solutions be determined by iterative schemes of computation?

What modes of local simplification of a problem are convenient/appropriate?

How can different solution techniques be compared and evaluated?

Distinguishing features of optimization as a mathematical discipline:

descriptive $\longrightarrow$ prescriptive

equations $\longrightarrow$ inequalities

linear/nonlinear $\longrightarrow$ convex/nonconvex

differential calculus $\longrightarrow$ subdifferential calculus

Finite-dimensional optimization: The case where a choice corresponds to selecting the values of a finite number of real variables, called *decision variables.* For general purposes the decision variables may be denoted by $x_1, \ldots, x_n$ and each possible choice therefore identified with a point $x = (x_1, \ldots, x_n)$ in the space $I\!R^n$. This is what we'll be focusing on in this course.

Feasible set: The subset C of $I\!R^n$ representing the allowable choices $x = (x_1, \ldots, x_n)$.

Objective function: The function $f_0(x) = f_0(x_1, \ldots, x_n)$ that is to be maximized or minimized over C.

Constraints: Side conditions that are used to specify the feasible set C within $I\!R^n$.

Equality constraints: Conditions of the form $f_i(x) = c_i$ for certain functions f_i on $I\!R^n$ and constants c_i in $I\!R^n$.

Inequality constraints: Conditions of the form $f_i(x) \leq c_i$ or $f_i(x) \geq c_i$ for certain functions f_i on $I\!R^n$ and constants c_i in $I\!R$.

Range constraints: Conditions restricting the values of some decision variables to lie within certain closed intervals of $I\!R$. Very important in many situations, for instance, are *nonnegativity constraints*: some variables x_j may only be allowed to take values ≥ 0; the interval then is $[0, \infty)$. Range constraints can also arise from the desire to keep a variable between certain upper and lower bounds.

Linear constraints: Range constraints or conditions of the form $f_i(x) = c_i$, $f_i(x) \leq c_i$, or $f_i(x) \geq c_i$, in which the function is *linear* in the standard sense of being expressible as sum of constant coefficients times the variables $x_1, \ldots, x_n$.

Data parameters: General problem statements usually involve not only decision variables but symbols designating known coefficients, constants, or other data elements. Conditions on such elements, such as the nonnegativity of a particular coefficient, are *not* among the "constraints" in a problem of optimization, since the numbers in question are supposed to be given and aren't subject to choice.

Mathematical programming: A traditional synonym for finite-dimensional optimization. This usage predates "computer programming," which actually arose from early attempts at solving optimization problems on computers. "Programming," with the meaning of optimization, survives in problem classifications such as linear programming, quadratic programming, convex programming, integer programming, etc.

EXAMPLE 1: Engineering Design

General description. In the design of some object, system or structure, the values of certain parameters can be chosen subject to some conditions expressing their ranges and interrelationships. The choice determines the values of a number of other variables on which the desirability of the end product depends, such as cost, weight, speed, bandwidth, reliability,.... Among the choices of the design parameters that meet certain performance specifications, which is the "best" by some criterion?

Particular case: optimal proportions of a can. A cylindrical can of a given volume V_0 is to be proportioned in such a way as to minimize the total cost of the material in a box of 12 cans, arranged in a 3×4 pattern. The cost expression takes the form $c_1 S_1 + c_2 S_2$, where S_1 is the surface area of the 12 cans and S_2 is the surface area of the box. (The coefficients c_1 and c_2 are positive.) A side requirement is that no dimension of the box can exceed a given amount D_0.

design parameters: $r =$ radius of can, $h =$ height of can

volume constraint: $\pi r^2 h = V_0$ (or $\geq V_0$, see below!)

surface area of cans: $S_1 = 12(2\pi r^2 + 2\pi r h) = 24\pi r(r+h)$

box dimensions: $8r \times 6r \times h$

surface area of box: $S_2 = 2(48r^2 + 8rh + 6rh) = 4r(24r + 7h)$

size constraints: $8r \leq D_0, \quad 6r \leq D_0, \quad h \leq D_0$

nonnegativity constraints: $r \geq 0, \quad h \geq 0$ (!)

Summary. The design choices that are available can be identified with the set C consisting of all the pairs $(r,h) \in \mathbb{R}^2$ that satisfy the conditions

$$r \geq 0, \quad h \geq 0, \quad 8r \leq D_0, \quad 6r \leq D_0, \quad h \leq D_0, \quad \pi r^2 h = V_0.$$

Over this set we wish to minimize the function

$$f_0(r,h) = c_1\big[24\pi r(r+h)\big] + c_2\big[4r(24r+7h)\big] = d_1 r^2 + d_2 rh,$$

where $d_1 = 24\pi c_1 + 96 c_2$ and $d_2 = 24\pi c_1 + 28 c_2$.

Comments. This example illustrates several features that are quite typically found in problems of optimization.

Redundant constraints: It is obvious that the condition $6r \leq D_0$ is implied by the other constraints and therefore could be dropped without affecting the problem. But in problems with many variables and constraints such redundancy may be hard to recognize. From a practical point of view, the elimination of redundant constraints could pose a challenge as serious as that of solving the optimization problem itself.

Inactive constraints: It could well be true that the optimal pair (r, h) (unique??) is such that either the condition $8r \leq D_0$ or the condition $h \leq D_0$ is satisfied as a strict inequality, or both. In that case the constraints in question are inactive in the local characterization of optimal point, although they do affect the shape of the set C. Again, however, there is little hope, in a problem with many variables and constraints, of determining by some preliminary procedure just which constraints will be active and which will not. This is the crux of the difficulty in many numerical approaches.

Redundant variables: It would be possible to solve the equation $\pi r^2 h = V_0$ for h in terms of r and thereby reduce the given problem to one in terms of just r, rather than (r, h). Fine—but besides being a technique that is usable only in special circumstances, the elimination of variables from (generally nonlinear) systems of equations is not necessarily helpful. There may be a trade-off between the lower dimensionality achieved in this way and other properties.

Inequalities versus equations: The constraint $\pi r^2 h = V_0$ could be written in the form $\pi r^2 h \geq V_0$ without affecting anything about the solution. This is because of the nature of the cost function; no pair (r, h) in the larger set C', obtained by substituting this weaker condition for the equation, can minimize f_0 unless actually $(r, h) \in C$. While it may seem instinctive to prefer the equation to the inequality in the formulation, the inequality turns to be superior in the present case because the set C' happens to be “convex,” whereas C isn't.

Convexity: This problem is not fully of “convex” type in itself, despite the preceding remark. Nonetheless, it can be made convex by a certain change of variables, as will be seen later. The lesson is that the formulation of a problem of optimization can be quite subtle, when it comes to bringing out crucial features like convexity.

EXAMPLE 2: Management of Systems

General description. A sequence of decisions must be made in discrete time which will affect the operation of some kind of "system," often of an economic nature. The decisions, each in terms of choosing the values of a number of variables, have to respect various limitations in resources. Typically the desire is to minimize cost, or maximize profit or efficiency, say, over a certain time horizon.

Particular case: an inventory model. A warehouse with total capacity a (in units of volume) is to be operated over time periods $t = 1, \ldots, T$ as the sole facility for the supply of a number of different commodities (or medicines, or equipment parts, etc.), indexed by $j = 1, \ldots, n$. The demand for commodity j during period t is the known amount $d_{tj} \geq 0$ (in volume units)—this is a *deterministic* approach to modeling the situation. In each period t it is possible not only to fill demands but to acquire additional supplies up to certain limits, so as to maintain stocks. The problem is to plan the pattern of acquiring supplies in such a way as to maximize the net profit over the T periods, relative to the original inventory amounts and the desired terminal inventory amounts.

inventory variables: x_{tj} units of j at the end of period t

inventory constraints: $x_{tj} \geq 0, \quad \sum_{j=1}^{n} x_{tj} \leq a$ for $t = 1, \ldots, T$

initial inventory: x_{0j} units of j given at the beginning

terminal constraints: $x_{Tj} = b_j$ (given amounts) for $j = 1, \ldots, n$

inventory costs: s_{tj} dollars per unit of j held from t to $t + 1$

supply variables: u_{tj} units of j acquired during period j

supply constraints: $0 \leq u_{tj} \leq a_{tj}$ (given availabilities)

supply costs: c_{tj} dollars per unit of j acquired during t

dynamical constraints: $x_{tj} = \max\big\{0,\, x_{t-1,j} + u_{tj} - d_{tj}\big\}$

rewards: p_{tj} dollars per unit of filled demand

filled demand: $\min\big\{\, d_{tj},\, x_{t-1,j} + u_{tj}\,\big\}$ units of j during period t

net profit: $\sum_{t=1}^{T} \sum_{j=1}^{n} \Big[p_{tj} \min\big\{\, d_{tj},\, x_{t-1,j} + u_{tj}\,\big\} - s_{tj} x_{tj} - c_{tj} u_{tj} \Big]$

Summary. The latter expression as a function of all the variables x_{tj} and u_{tj} for $t = 1, \ldots, T$ and $j = 1, \ldots, n$ is to be maximized subject to the inventory constraints, terminal constraints, supply constraints and the dynamical constraints. (These constraints can be viewed as determining a certain subset of $\mathbb{R}^{2Tn}$.)

Comments. Again, there are many insights from this example into the challenges that must be faced in optimization theory and practice.

Large-scale context: The number of variables and constraints that can be involved in a problem may well be very large, and the interrelationships may be too complex to appreciate in any direct manner. This calls for new ways of thinking and for more reliance on guidelines provided by theory.

Uncertainty: Clearly, the assumption that the demands d_{tj} are known precisely in advance is unrealistic for many applications, although by solving the problem in this case one might nonetheless learn a lot. To pass from deterministic modeling to *stochastic* modeling, where each d_{tj} is a random variable (and the same perhaps for other data elements like a_{tj}), it is necessary to expand the conceptual horizons considerably. The decision vector $(u_{t1}, \ldots, u_{tn})$ at time t must be viewed as an *unknown function* of the "information" available to the decision maker at that time, rather than just at the initial time.

Dependent variables: The values of the variables x_{tj} are completely determined by the values of the variables u_{tj} for $t = 1, \ldots, T$ and $j = 1, \ldots, n$ through the dynamical equations and the initial values. In principal, therefore, some specific expression in the latter variables could be substituted for each x_{tj}, and the dimensionality of the problem could thereby be cut in half. But this trick, because it hides basic aspects of structure, could actually make the problem harder to analyze and solve.

Constraints versus penalties: The requirements that $\sum_{j=1}^{n} x_{tj} \leq a$ for $t = 1, \ldots, T$ and $x_{Tj} = b_j$, although harmless-looking, are potentially troublesome in their effect on computation and even on the existence and analysis of solutions. Better modeling would involve some recourse in the eventuality of these conditions not being satisfied. For instance, instead of a constraint involving the capacity one could incorporate into the function being minimized a penalty term, which kicks in when the total amount being stored rises above a (perhaps with the interpretation that extra storage space has to be rented).

Max and min operations: The "max" operation in the dynamical constraints and the "min" operation in the expression of the net profit force the consideration of functions and mappings that can't be handled by ordinary calculus. Sometimes this is unavoidable and points to the need for fresh developments in analysis. Other times it is an unnecessary artifact of the formulation. The present example fits with the latter. Really, it would be better to introduce still more variables: v_{tj} as the amount of good j used to meet demands at time t. In terms of these variables, constrained by $0 \leq v_{tj} \leq d_{tj}$, the dynamics would take the linear form

$$x_{tj} = x_{t-1,j} + u_{tj} - v_{tj}$$

and the profit expression would likewise be linear:

$$\sum_{t=1}^{T} \sum_{j=1}^{n} \Big[p_{tj} v_{tj} - s_{tj} x_{tj} - c_{tj} u_{tj} \Big].$$

Hidden assumptions: The alternative model just described with variables v_{tj} is better in other ways too. The original model had the hidden assumption that demands in any period should always be met as far as possible from the stocks on hand. But this might be disadvantageous if rewards will soon be higher, and inventory can only be built up slowly due to the constraints on availability. The alternative model allows sales to be held off in such circumstances.

EXAMPLE 3: Identification of Parameters

General description. A mathematical model has been formulated for a given situation, but to implement it the values of a number of parameters must be specified. A body of data is known through experiment or observation. The task is to determine the parameter values that best fit the data. Here, in speaking of the "best" fit, reference is evidently being made to some criterion for optimization, but there isn't just one interpretation always of which criterion to use. (Note a linguistic pitfall: "the" best fit suggests uniqueness of the answer being sought, but even relative to a single criterion there could be more than one choice of the parameters that is optimal.) Applications are found in statistics (regression, maximum likelihood), econometrics, and virtually every area of science.

Particular case: "least squares" estimates. Starting out very simply, suppose that two variables x and y are being modeled as related by a linear law $y = ax + b$, either for inherent theoretical reasons or as a first-level approximation. The values of a and b are not known *a priori* but must be determined from the data, consisting of a large collection of pairs $(x_k, y_k) \in \mathbb{R}^2$ for $k = 1, \ldots, N$. These pairs have been gleaned from experiments (where random errors of measurement could arise along with other discrepancies due to oversimplifications in the model). The error expression

$$E(a,b) = \sum_{k=1}^{N} \left| y_k - (ax_k + b) \right|^2$$

is often taken as representing the goodness of the fit of the parameter pair (a,b). The problem is to minimize this over all $(a,b) \in I\!R^2$. More generally, instead of a real variable x and a real variable y one could be dealing with a vector $x \in I\!R^n$ and a vector $y \in I\!R^m$, which are supposed to be related by a formula $y = Ax + b$ for a matrix $A \in I\!R^{m \times n}$ and a vector $b \in I\!R^m$. Then the error expression $E(A,b)$ would depend on the $m \times (n+1)$ components of A and b.

Comments. This kind of optimization is entirely technical: the introduction of something to be optimized is just a mathematical construct. Still, in analysis and computation of solutions the same challenges arise as in other settings.

Constraints: The problem, as stated so far, concerns the unconstrained minimization of a certain quadratic function in the parameters, but it is easy to imagine situations where the parameters may be subject to various side conditions. In the case of $y = ax + b$, for instance, it may be known on the basis of theory for the variables in question that $1/2 \leq a \leq 3/2$, while $b \geq -1$. In the multidimensional case of $y = Ax + b$ there may be the requirement of A being symmetric (with $m = n$), which would entail the imposition of $n(n-1)/2$ linear constraints of the form $a_{ji} - a_{ij} = 0$. Perhaps for some reason one also needs to have $a_{11} \geq a_{22} \geq \cdots \geq a_{nn}$, and so forth. (In the applications that are made of least squares estimation, such conditions are often neglected, and the numerical answer obtained is simply "fixed up" if it doesn't have the right form. But this is clearly not good methodology.)

Nonlinear version: A so-called problem of *linear* least squares has been presented, but the same ideas can be used when the underlying relation between x and y is supposed to be nonlinear. For instance, a law of the form $y = e^{ax} - e^{bx}$ would lead to an error expression

$$E(a,b) = \sum_{k=1}^{N} \left| y_k - (e^{ax_k} - e^{bx_k}) \right|^2.$$

In minimizing this with respect to $(a,b) \in I\!R^2$, we would not be dealing with a quadratic function, but something much more complicated. The graph of E in a problem of nonlinear least squares could have lots of "bumps" and "dips," which could make it hard to find the minimum computationally.

Beyond squares: Many other expressions for error could be considered instead of a sum of squares, and in some situations one of these might be preferable. Back in the elementary case of $y = ax + b$, for example, one could look at

$$E(a,b) = \sum_{k=1}^{N} \big| y_k - (ax_k + b) \big|.$$

A different (a,b) would then be "optimal." The optimization problem would have a technically different character as well, because E would not only fail to be quadratic but would even fail to be differentiable (at points (a,b) where $y_k - (ax_k + b) = 0$ for some k). Still more dramatic in this regard, yet quite justifiable as an approach, would be the error expression

$$E(a,b) = \max_{k=1,\ldots,N} \big| y_k - (ax_k + b) \big|.$$

The formula in this case means that the value assigned by E to the pair (a,b) is the largest value occurring among the errors $|y_k - (ax_k + b)|$, $k = 1, \ldots, N$. It is this maximum deviation that we wish to make as small as possible. Once more, E is not a differentiable function on $\mathbb{R}^2$, although it nevertheless has plenty of good structure that can be utilized in analyzing and solving the problem of optimization.

Other problems of approximation: Similar in character are problems in which a given function $f : [c,d] \mapsto \mathbb{R}$ is to be approximated by a linear combination of elementary functions $f_j : [c,d] \mapsto \mathbb{R}$, $j = 1, \ldots, n$. The expressions $f_j(t)$ could be monomials like t^j (in which case f is being approximated by a polynomial of a certain degree or less), or on the other hand they could be sine and cosine terms (in which case f is being approximated by a truncated Fourier series), say. With the unknown coefficient of f_j denoted by s_j, the problem would be to choose the vector $s = (s_1, \ldots, s_n) \in \mathbb{R}^n$ so as to minimize an error expression like

$$E_p(s) = \int_c^d \big| f(t) - \textstyle\sum_{j=1}^{n} s_j f_j(t) \big|^p dt, \text{ with } p \in [1, \infty)$$

(approximation in norm of $\mathcal{L}^p[c,d]$, the case of $p = 2$ being another version of "least squares"), or alternatively

$$E_\infty(s) = \max_{c \le t \le d} \big| f(t) - \textstyle\sum_{j=1}^{n} s_j f_j(t) \big|$$

(this case, in the space $\mathcal{L}^\infty[c,d]$, being termed *Chebyshev* approximation).

EXAMPLE 4: Variational Principles

General description. The linear and nonlinear equations that are the focus of much of numerical analysis are often associated in hidden ways with problems of optimization. For an equation of the form $F(x) = 0$, involving a mapping $F : \mathbb{R}^n \mapsto \mathbb{R}^n$, a *variational principle* is an expression of F as the gradient mapping ∇f associated with some function $f : \mathbb{R}^n \mapsto \mathbb{R}$. Such an expression leads to the interpretation that the desired x satisfies a first-order optimality condition with respect to f. Under certain additional conditions on F, it may even be concluded that x minimizes f, at least "locally." A route to solving $F(x) = 0$ is thereby opened up in terms of minimizing f. Quite similar in concept are numerous examples where instead of solving an equation $F(x) = 0$ with $x \in \mathbb{R}^n$ one is interested in solving $A(u) = 0$ where u is some unknown *function*, and A is a mapping from a function space (e.g. a certain Hilbert space) into itself. In particular, A might be a differential operator, so that an ordinary or partial differential equation is at issue. A variational principle then characterizes the desired u as providing the minimum, say, of some functional on the space. In fact, many of the most famous differential equations of physics have such an interpretation, including Newton's laws of motion (the local variational principle of "least action"). On a different front, one can think of conditions of price equilibrium in economics that can be characterized as stemming from the actions of a multitude of "economic agents," like producers and consumers, all optimizing from their individual perspectives. Yet again, the equilibrium state following the reactions which take place in a complicated chemical brew may be characterized through a variational principle as the configuration of substances that minimizes a certain energy function.

Particular case: the Dirichlet problem. A classical problem in PDE's, posed in its most elementary form, concerns an unknown function $u(y_1, y_2)$ on a closed, bounded region $\Omega \subset \mathbb{R}^2$ with boundary curve Γ. The function, assumed to be continuous on Ω and twice differentiable on the interior of Ω, is required to satisfy, in terms of a given function φ on Ω, the partial differential equation

$$\frac{\partial^2 u}{\partial y_1^2}(y_1, y_2) + \frac{\partial^2 u}{\partial y_2^2}(y_1, y_2) = \varphi(y_1, y_2)$$

inside Ω as well as the boundary condition

$$u(y_1, y_2) \equiv 0 \text{ on } \Gamma.$$

It turns out that the solution to this problem is the unique function that minimizes the expression

$$J(u) = \iint_{\Omega} \left[\varphi(y_1, y_2) u(y_1, y_2) + \frac{1}{2} \frac{\partial u}{\partial y_1}(y_1, y_2)^2 + \frac{1}{2} \frac{\partial u}{\partial y_2}(y_1, y_2)^2 \right] dy_1 dy_2$$

over all functions u satisfying the boundary condition. Although this is not a problem of *finite-dimensional* optimization, because u ranges over a space with "infinitely many degrees of freedom," one does get such a problem in passing to a *discretized* version or an approximate problem in which u is restricted to be a linear combination of a certain collection of basic functions (some kind of truncated series), as must inevitably be done in bringing numerical methods to bear. It is obvious that this is another way that optimization problems of very high dimension could arise. A major branch of theory has to address the question of how an infinite-dimensional problem can be approximated better and better by a sequence of finite-dimensional problems, what kind of convergence can be obtained from the respective solutions, and so on.

Comments. In the study of variational principles, optimization theory can provide interesting insights quite independently of whether a numerical solution to a particular case is sought or not.

Classical roots: Optimization problems over function spaces have been studied since the 17th century, and they have been very influential not only in the discovery of variational principles but in the development of tools of analysis, especially functional analysis and topology. This branch of the subject has traditionally been referred to by the quaint title of *the calculus of variations.* A closely related modern counterpart, to be discussed in the next example, is the theory of optimal control.

Unilateral constraints: While the side conditions considered with classical PDE's are typically equations of some kind, many problems handled nowadays involve inequalities. For instance, it may be required above that

$$a(y_1, y_2) \leq u(y_1, y_2) \leq b(y_1, y_2)$$

for certain functions a and b given on Ω. From the standpoint of the PDE, it's not completely clear what this is supposed to mean, but in the context of the optimization of $J(u)$ the meaning is evident: this condition, in addition to the boundary condition already imposed, is to restrict further the set C of functions u over which the minimization is to be carried out. From the theory of such an optimization problem, a characterization of the minimizing u can be derived. This characterization is in the form not of a PDE as such, but a sort of generalized PDE called a *variational inequality*. The point is that optimization theory, through notions of variational principles, can provide guidelines to the appropriate generalization of PDE's where a direct approach, in terms of just playing with the equation, might go astray.

EXAMPLE 5: Optimal Control

General description: The evolution of a system in continuous time t can often be characterized by an ordinary differential equation $\dot{x}(t) = f(t, x(t))$ with $x(0) = x_0$ (initial condition), where $\dot{x}(t) = (dx/dt)(t)$. (Equations involving higher derivatives are typically reducible to ones of first order by well known tricks.) Here $x(t)$, called the *state* of the system, is a point in $\mathbb{R}^n$. This is *descriptive* mathematics. We get *prescriptive* mathematics when the ODE involves parameters for which the values can be chosen as a function of time: $\dot{x}(t) = f(t, x(t), u(t))$, where $u(t) \in U \subset \mathbb{R}^m$. Without going into the mathematical details necessary to provide a rigorous foundation, the idea can be appreciated that under various assumptions there will be a mapping which assigns to each choice of a *control* function $u(\cdot)$ over a time interval $[0, T]$ a corresponding state trajectory $x(\cdot)$. Then, subject to whatever restrictions may be necessary or desirable on these functions, one can seek the choice of $u(\cdot)$ which is optimal according to some criterion. Although such a problem would be infinite-dimensional, a finite-dimensional version would arise as soon as the differential equation is approximated by a difference equation in discrete time.

Particular case: an expedition to the moon. A space ship is to be sent from a certain location on earth to land in an area on the moon. The state of the ship at any time t can be described by a finite collection of values $x_1(t), \ldots, x_n(t)$, which may include position coordinates, velocity coordinates, and other aspects of deployment in flight related to orientation, spin, etc. The state is thus a vector $x(t) \in \mathbb{R}^n$. Without intervention, the ship would move in a deterministic manner according to some ODE dictated by the laws of mechanics, but influence can be exerted through the firing of various rockets incorporated in the ship, for which there is some amount of control over the angle and thrust. To look at this figuratively, imagine a control panel with m levers which can be set at positions between 0 and 1. A point $u(t) = (u_1(t), \ldots, u_m(t))$ in the cube $U = [0, 1]^m \subset \mathbb{R}^m$

specifies the positions of all the levers. Once a function $u : [0, T] \mapsto U$ has been chosen, the trajectory x followed by the space ship will be completely determined over the time interval $[0, T]$ as the solution to an ODE $\dot{x}(t) = f(t, x(t), u(t))$ with $x(0) = x_0$. Restrict attention now to the class of control functions u such that the final state $x(T)$ has its position coordinates in the targeted area of the moon, its velocity coordinates all 0, and so forth; this will be a constraint of the form $x(T) \in E$. Further make restrictions like $x(t) \in X(t) \subset I\!R^n$, for instance to ensure that the trajectory of the ship does not penetrate the earth or the moon. Over the control functions so described, the problem is to minimize some expression like

$$J(u) = \int_0^T g(t, x(t), u(t))dt$$

giving the cost of the control, say. Other possibilities include looking for a control function that gets the ship to its destination in the least time, for instance.

Comments.

Control in discrete time: Very similar problems can be set up in discrete rather than continuous time, not just as an expedient for numerical purposes, but as appropriate models in themselves. Such problems are finite-dimensional. A case in point is the inventory problem in Example 2, where the x_{tj}'s are state variables and the u_{tj}'s are control variables.

Stochastic version: The system being guided may be subject to random disturbances, which the controller must react to. Further, there may be difficulty in knowing exactly what the state is at any time t, due to the shortcomings of sensors and measurement errors (another random effect). Control must then be framed in terms of mappings which give the response at time t that is most appropriate to the particular information available right then about $x(t)$. This is the formidable subject of *stochastic optimal control*, which at present is only able to cope with rather special cases.

Adaptive version: Also very interesting as a mathematical challenge, but largely out of reach of current concepts and techniques, is *adaptive control*, where the controller has not only to react to unexpected events but learn the basics of the system being controlled as time goes on. This is a bit like getting behind the wheel of a car in bad weather when the roads are icy. In choosing the control function, the desire to arrive at the destination in the quickest manner compatible with the configuration of the roads and hills may have to be compromised with time spent on "test skids" to see what the tires can take. A major difficulty in this area is the clear formulation of the objective in the optimization, as well as the identification of what can or can't be assumed about the imperfectly known system.

Control of PDE's: The state of a system may be given by an element of a function space rather than a point in $\mathbb{R}^n$, as when the problem revolves around the temperature distribution at time t over a solid body represented by a closed, bounded region $\Omega \subset \mathbb{R}^3$. The temperature can be influenced by heating or cooling elements arrayed on the surface of the body. How should these elements be operated in order to bring the temperature of the body uniformly within a certain range—in the shortest possible time, or with the least expenditure of energy?

EXAMPLE 6: Optimal Scheduling

General description. Choices must be made about the order in which certain actions ought to be taken, as well as scope of the actions. Decisions may concern not only the values of continuous variables but *discrete* variables, which can take on only integer values, or even *logical* variables, which are limited to 0 and 1. There may thus be a mixture of finite-dimensional optimization and *combinatorial* optimization. Many such problems are almost intractable, even when posed in just a halfway realistic form, but there are notable exceptions.

Particular case: flight scheduling. An airline must set up its weekly schedule of flights. This involves specifying not only the departure and arrival times but the numbers of flights between various destinations (these numbers have to be treated as integer variables). Constraints involve, among other things, the availability of aircraft and crew and are greatly complicated by the need to follow what happens to each individual plane and crew member. A particular plane, having flown from Seattle to New York, must next take off from New York, and it can't do so without a certified pilot, who in the meantime has arrived from Atlanta and gotten the right amount of rest, and so on. Aircraft maintenance requirements are another serious issue along with the working requirements of personnel based in different locations and having to return home at specified intervals. The flight schedule must obviously take into account the passenger demand for various routes and times, and whether they are nonstop. To the important extent that random variables are involved, not only in the demands but in the possibility of mechanical breakdowns, sick crew members and weather delays, various recourses and penalties must be built into the model. Somewhere in this picture there should be an approach to scheduling which optimizes relative to cost or profit, say, but the example illustrates the great difficulties that can arise in formulating mathematically the appropriate objective as well as the constraints.

Comments. Successful solution of problems in this vein depends usually on a marriage between techniques in ordinary "continuous" optimization and special ways of handling certain kinds of combinatorial structure. One of the main areas in which there have been strong achievements and interesting theoretical developments is that of *networks* (directed graphs). Remarkably many problems can be set up and solved in terms of network flows and potentials.

Overview of areas of optimization: Because this course will concentrate on finite-dimensional optimization (as in Examples 1, 2, and 3), infinite-dimensional applications to variational principles (as in Example 4) and optimal control (as in Example 5) will not be covered directly, nor will special tools for applications involving integer variables or combinatorial optimization (as in Example 6) be developed. Still, many of the ideas that will be dealt with here are highly relevant as background for such other areas of optimization. In particular, infinite-dimensional problems are often approximated by finite-dimensional ones through some kind of discretization of time, space, or probability.