Supplementary material on optimization for the course Portfolio Theory and Risk Evaluation

September 4, 2002

Optimization

Optimization is about the abstract problem of finding a minimum point, x, of a real valued function, f, in some set X,

$$\min_{x \in X} f(x).$$

We refer to the function f as the objective function and the set X as the feasible set. This note is intended to give the fundamentals of optimization in a concise manner. It does not cover numerical methods to solve different kind of optimization problems. For this we refer the reader to a course on optimization or numerical analyis. We will only consider the finite-dimensional case, that is when X is some subset of $\mathbb{R}^n$.

Observe, that problems that naturally come in maximization form, for example profit maximization, can be converted into a minimization problem by changing the sign of the objective.

Optimality

A point $\bar{x} \in \mathbb{R}^n$ is said to be a global minimum of a function $f : \mathbb{R}^n \to \mathbb{R}$, if $f(x) \geq f(\bar{x})$ for all $x \in \mathbb{R}^n$. A point $\bar{x}$ is said to be a local minimum if $f(x) \geq f(\bar{x})$ for all x in a neighborhood of $\bar{x}$, i.e. for all x such that $\|x - \bar{x}\| \leq \epsilon$ for some $\epsilon > 0$.

Observe that a minimization problem need not have optimal solutions, local or global, for example $\min_{x \geq 0} e^{-x}$ does not have a solution.

Convexity

An important concept in optimization is convexity because it makes all local optima global.

A set $C \subset \mathbb{R}^n$ is said to be convex if for all $x, y \in C$ and $\alpha \in [0,1]$ it holds that $\alpha x + (1-\alpha)y \in C$.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if the function lies below its cords, i.e. $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$ for $\alpha \in [0,1]$. The function f is said to be strictly convex if the previous inequality is strict for $\alpha \in (0,1)$.

If f is at least twice continuously differentiable one can show that the following three statements are equivalent. Here $\nabla f(x)$ denotes the gradient of f at x, that is the vector of partial derivatives, and $\nabla^2 f(x)$ denotes the Hessian of f at x, that is the matrix of second order partial derivatives.

(i) f is convex on the whole $\mathbb{R}^n$.

(ii) $f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x})$ for all $\bar{x}$ and $x \in \mathbb{R}^n$ (i.e. the function lies above its tangents).

(iii) $\nabla^2 f(x)$ is positive semidefinite ($h^T \nabla^2 f(x) h \geq 0$ for all $h \in \mathbb{R}^n$) for all $x \in \mathbb{R}^n$.

It follows from (ii) that if f is convex and $\nabla f(\bar{x}) = 0$ then $\bar{x}$ is a *global* minimum.

Taylor

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and assume that you have calculated $f(\bar{x})$, $\nabla f(\bar{x})$ and $\nabla^2 f(\bar{x})$ at a given point $\bar{x}$. Then f can be approximated in a neighborhood of $\bar{x}$ in the following manner:

$$f(x) = f(\bar{x} + d) = f(\bar{x}) + \nabla f(\bar{x})^T d + \frac{1}{2} d^T \nabla^2 f(\bar{x}) d + o(\|d\|^2)$$

where $o(\|d\|^2)/\|d\|^2$ goes to zero when $\|d\|$ goes to zero.

Using this one can show that if $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is positive definite then $\bar{x}$ is a local minimum of f.

Karush-Kuhn-Tucker conditions (KKT-conditions)

Consider a non-linear optimization problem on the following form.

$$(\mathrm{P}^{\leq}) \quad \begin{array}{ll} \min & f(x) \\ \text{subject to} & g_i(x) \leq b_i, \quad i = 1, \ldots, m, \\ & x \in \mathbb{R}^n. \end{array}$$

Definition $\bar{x} \in \mathbb{R}^n$ is a KKT-point for $(\mathrm{P}^{\leq})$ if there exist real scalars $\lambda_1, \ldots, \lambda_m$ that together with $\bar{x}$ satisfy the following conditions:

(i) $\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0$ (n conditions)

(ii) $g_i(x) \leq b_i$, $i = 1, \ldots, m$

(iii) $\lambda_i \geq 0$, $i = 1, \ldots, m$

(iv) $\lambda_i(g_i(\bar{x}) - b_i) = 0, i = 1, \ldots, m$

The KKT-conditions above are normally necessary conditions for $\bar{x}$ to be a local minimum to ($\mathrm{P}^{\leq}$) (i.e. $\bar{x}$ is a local minimum implies that $\bar{x}$ is a KKT-point). For this to be true the problem ($\mathrm{P}^{\leq}$) need to fulfill some sort of "Constraint Qualification" (CQ). Two such (CQ) are

(i) all g_i in ($\mathrm{P}^{\leq}$) are linear, or

(ii) all g_i are convex functions and there exists at least one point x such that $g_i(x) < b_i$ for all i.

For convex problems that is problems where f and all the g_i's are convex functions the KKT-conditions above are sufficient, i.e. $\bar{x}$ is a KKT-point implies that $\bar{x}$ is a global minimum point..

Lagrange conditions

Consider now a problem with only equalities instead of inequalities, i.e. a problem on the following form.

$$(\mathrm{P}^{=}) \quad \begin{array}{ll} \min & f(x) \\ \text{subject to} & g_i(x) = b_i, \quad i = 1, \ldots, m, \\ & x \in \mathbb{R}^n. \end{array}$$

In this case the KKT-conditions are simplier and one usually call them Lagrange conditions.

Definition $\bar{x} \in \mathbb{R}^n$ satisfy the Lagrange conditions ($\mathrm{P}^{=}$) if there exist real scalars $\lambda_1, \ldots, \lambda_m$ which together with $\bar{x}$ satisfy the following conditions:

(i) $\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0$ (n conditions)

(ii) $g_i(x) = b_i$, $i = 1, \ldots, m$.

If $\bar{x}$ is a local optimal solution to ($\mathrm{P}^{=}$), and ($\mathrm{P}^{=}$) satisfy a constraint qualification, then $\bar{x}$ satisfy the Lagrange conditions of ($\mathrm{P}^{=}$). Two examples of constraint qualifications to ($\mathrm{P}^{=}$) are

(i) all g_i in ($\mathrm{P}^{\leq}$) are linear, or

(ii) the m gradients $\nabla g_1(x), \nabla g_2(x), \ldots, \nabla g_m(x)$ are linearly independent for all x.

The Lagrange function of ($\mathrm{P}^=$) is defined as follows

$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i (g_i(x) - b_i).$$

Using this notation one can write the Lagrange conditions on the following form:

(i) $\frac{\partial L(\bar{x},\bar{\lambda})}{\partial x_j} = 0$, $j = 1, \ldots, n$,

(ii) $\frac{\partial L(\bar{x},\bar{\lambda})}{\partial \lambda_i} = 0$, $i = 1, \ldots, m$.

Note, however, that $(\bar{x}, \bar{\lambda})$ is normally not a minimum nor a maximum point of L, but typically a saddle point.

For certain types of problems the Lagrange conditions are also sufficient optimality conditions. One such problem type is when the functions g_i are all linear functions and f is a convex differentiable function.

Duality and Everett's theorem

Given the problem

$$(\mathrm{P}) \quad \begin{array}{ll} \min & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & x \in X \end{array}$$

one say that for fixed λ the problem

$$(\mathrm{P}_\lambda) \quad \begin{array}{rl} \min & f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \\ \text{subject to} & x \in X \end{array}$$

is a relaxation of (P) if $\lambda_i \geq 0$.

Let us denote the optimal objective value of (P_λ) by $\phi(\lambda)$. Then it is easy to verify that $\phi(\lambda) \leq f(x)$ for all x feasible to (P). This means that $\phi(\lambda)$ provides a lower bound on how well it is possible optimize f. It is then natural to try to make this bound as tight as possible, that is to maximize $\phi(\lambda)$. This is the so called Lagrangean dual problem of (P):

$$(\mathrm{D}) \quad \begin{array}{ll} \max & \phi(\lambda) \\ \text{subject to} & \lambda \geq 0. \end{array}$$

Relating to this we have the following simple but powerful theorem.

Theorem (Everett) Suppose that $x(\lambda)$ solves (P_λ) then $x(\lambda)$ also solves the problem (P) but where the zeros of right-hand side of the constraints have been replaced by $g_i(x(\lambda))$.

Quadratic functions

A function $p : \mathbb{R}^n \to \mathbb{R}$ is said to be quadratic if it can be written on the form

$$p(x) = \frac{1}{2}x^T Q x + c^T x + c_0 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}x_i x_j + \sum_{i=1}^{n} c_j x_j + c_0,$$

where Q is a symmetric real-valued $n \times n$-matrix, $c \in \mathbb{R}^n$ and $c_0 \in \mathbb{R}$.

The gradient of p can be written as

$$\nabla p(x) = Qx + c$$

and the Hessian

$$\nabla^2 p(x) = Q.$$

It is easy to show that p is strictly convex if and only if Q is positive definite ($h^T Q h > 0$ for $h \neq 0$) , and that p is convex if Q is positive semidefinite ($h^T Q h \geq 0$ for all h).

Assume that $\nabla p(\bar{x}) = 0$, that is $Q\bar{x} = -c$ then simple calculations show that $p(x) = p(\bar{x}) + \frac{1}{2}(x - \bar{x})^T Q(x - \bar{x})$ for all $x \in \mathbb{R}^n$. Now, if Q is positive definite then apparently $p(x) > p(\bar{x})$ for all $x \neq \bar{x}$ which implies that $\bar{x}$ is a *global* minimum point to p.

Quadratic problems and their optimimality conditions

If the objective is a quadratic function and the the constraints are linear, then the problem is called a QP-problem:

$$\begin{array}{lll} & \min & \frac{1}{2}x^T Q x + c^T x \\ (\text{QP}) & \text{subject to} & a_i^T x = b_i, \quad i = 1, \dots, m \\ & & x_j \geq 0, \qquad j = 1, \dots, n. \end{array}$$

Here, we have the constraints on so called standard form, i.e. all the variables need to be non-negative and the linear constraints are equalities. Using matrix-notation, we may write the feasible set as $\{x \mid Ax = b, x \geq 0\}$.

A point $\bar{x}$ is said to be a KKT-point of (QP), if there exist multipliers λ_i and $\mu_j \geq 0$ such that

(i) $Q\bar{x} + c + \sum_{i=1}^{m} \lambda_i a_i - \mu = 0$ (n conditions)

(ii) $a_i^T \bar{x} = b_i$, $i = 1, \dots, m$

(iii) $\mu_j \geq 0$, $j = 1, \dots, n$

(iv) $\mu_j \bar{x}_j = 0, j = 1, \dots, n$

If all the variables x_j in (QP) are free to take any value then the KKT conditions simplify to the Lagrange conditions that there exist λ_i such that

(i) $Q\bar{x} + c + \sum_{i=1}^{m} \lambda_i a_i = 0$ (n conditions)

(ii) $a_i^T \bar{x} = b_i$, $i = 1, \ldots, m$.

Note that finding such $\bar{x}$ and λ amounts to solving a linear system of equations with $m + n$ unknowns and the same number of equations.

We have the following theorem.

Theorem Suppose that Q is positive semidefinite then $\bar{x}$ is a KKT-point if and only if $\bar{x}$ is an optimal solution to (QP).

Linear programming

A very simple but useful problem class is linear programming. A linear program on so called standard form is

$$\text{(LP)} \quad \begin{array}{lll} \min & \sum_{j=1}^{n} c_j x_j & \\ \text{subject to} & \sum_{j=1}^{n} a_{ij} x_j = b_i, & i = 1, \ldots, m \\ & x_j \geq 0, & j = 1, \ldots, n. \end{array}$$

or using matrix notation

$$\text{(LP)} \quad \begin{array}{ll} \min & c^T x \\ \text{subject to} & Ax = b, \\ & x \geq 0. \end{array}$$

What are the KKT conditions for this problem?

Linear programming is important for several reasons. First, and perhaps most important is that many decisions problems in the real world, lend themselves to be modelled very well by linear programs. Second, all convex problems can be approximated to arbitrary accuracy using linear programming. A third reason is that there exist very good methods to solve linear programming problems, and problems of huge dimensions can be solved. For example, in so called crew scheduling problems it is not unusual that airlines on a routine basis solve problems with millions of variables (n) and several thousand constraints (m). Also in portfolio analysis some extraordinary large linear programming models have been formulated and successfully solved and applied for financial decision making.