

Preliminary

This course is strongly based on the monography on Operations Research by Carter, Price and Rabadi [1], and in material obtained from different sources (quoted when needed through the slides).

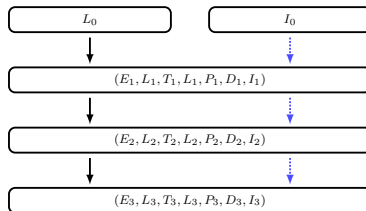
Nonlinear optimization: Summary

- 1 Nonlinear optimization
- 2 Concavity/convexity
- 3 Iterative methods
- 4 Constrained optimization
 - Lagrange multipliers
 - Karush-Kuhn-Tucker conditions

Example

Production cost

A company aims at minimizing the production cost during a series of production periods of time T_i , characterised by the demand D_i , equipment capacities and material limitations E_i , labor force L_i (which cost depends quadratically on the difference of labor force between two periods $C_L(L_i - L_{i-1})^2$), productivity of each worker P_i , the number of units of inventory at the end of each production period I_i and the cost to bring them to the next production period C_I .



Example

The problem, as stated, aims at minimizing the function:

$$f(\vec{L}, \vec{l}) = \sum_{i=1}^T C_L(L_i - L_{i-1})^2 + C_l l_i$$

subject to:

$$\begin{cases} L_i P_i \leq E_i \\ l_{i-1} + L_i P_i \geq D_i \\ l_i = l_{i-1} + L_i P_i - D_i \\ L_i, l_i \geq 0, \forall i = 1, \dots, T \end{cases}$$

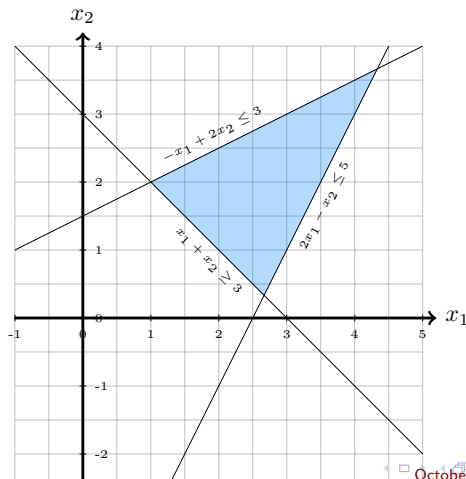
Quadratic objective function with linear constraints.

Introduction

- We want to obtain the best solution to a mathematical programming problem in which both objective function and constraints have general non-linear forms.
- Most of the problems are non-linear indeed!
 - Unconstrained Problems (often dealt with differential calculus)
 - Constrained Problems (may include systems of equations to be solved)
- Classical underlying mathematical theories do not necessarily provide practical methods suitable for efficient numerical computation.
- Points of optimality can be anywhere inside the problem boundaries
- No methods applicable to all non-linear problems

Feasible region

Set of points satisfying all the constraints (the area between constraint boundaries).



Global and local extremes



A local maximum of the function $f(x)$ exists in x^* if there is a small positive number ϵ such that

$$f(x^*) > f(x), \forall x \in \mathbf{R} : \|x - x^*\| < \epsilon$$

A global maximum of $f(x)$ exists in x^* if

$$f(x^*) > f(x), \forall x \in \mathbf{R}$$

(analogous definitions for local/global minimum)

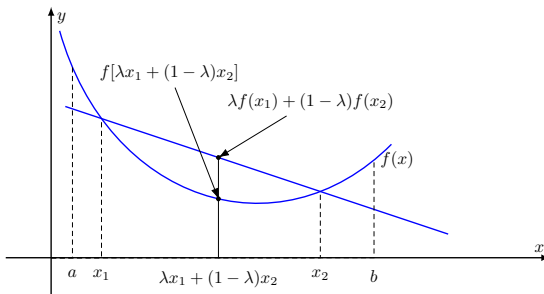
Concavity/convexity: Summary

- 1 Nonlinear optimization
- 2 Concavity/convexity
- 3 Iterative methods
- 4 Constrained optimization
 - Lagrange multipliers
 - Karush-Kuhn-Tucker conditions

For a continuous convex function, given any two points x_1 and x_2 :

$$f[\lambda x_1 + (1 - \lambda)x_2] \leq \lambda f(x_1) + (1 - \lambda)f(x_2), \quad 0 \leq \lambda \leq 1 \quad (1)$$

(analogous situation for a continuous concave function)



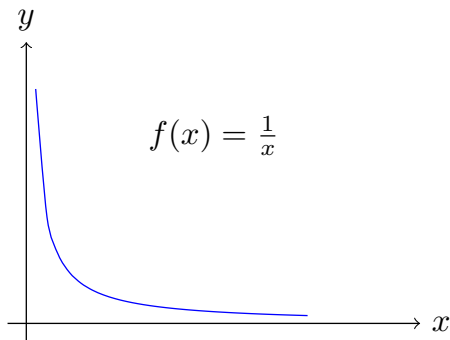
Convexity

The term "convex" can be applied both to sets and functions. A set $S \in \mathbf{R}^n$ is a *convex set* if the straight line segment connecting any two points in S lies entirely inside S .

Note that $f(x)$ is a convex function if its domain S is a convex set and if for any two points x_1 and x_2 in S Eq. 1 holds.

Exercise 1 Can you draw a convex function that depends on two variables, $f(x, y)$? Can you generalize Eq. 1?

- If a convex nonlinear function is to be optimized without constraints, a global minimum may occur when $f'(x) = 0$. Not always this is the case:



- If the feasible region for a nonlinear programming problem is convex, each of the constraint functions is convex and are of the form $g_i(x) \leq b_i$

- A local minimum is guaranteed to be a global minimum for a convex objective function in a convex feasible region, and
- a local maximum is guaranteed to be a global maximum for a concave objective function in a convex feasible region.

Many functions in nonlinear programming problems are neither concave nor convex!

Exercise 2 Is the function $f(x) = x^2$ convex? Are the regions defined by $x^2 = 4$ or $x^2 \geq 9$ convex?

Exercise 3 Are the functions $f(x, y) = 3x^2 - 2xy + y^2 + 3e^{-x}$ and $g(x, y) = x^4 - 8x^3 + 24x^2 - 32x + 16$ convex?

Exercise 4 Consider these two nonlinear problems[1]:

$$\begin{array}{l|l}
 \text{minimize} & f(x, y) = x - 2xy + 2y \\
 \text{subject to} & \begin{cases} x^2 + 3y^2 \leq 10 \\ 3x + 2y \geq 1 \\ x, y \geq 0 \end{cases}
 \end{array}
 \quad \Bigg| \quad
 \begin{array}{l|l}
 \text{minimize} & f(x, y) = x - 2xy + 2y \\
 \text{subject to} & \begin{cases} x^3 - 12x - y \geq 0 \\ x \geq 1 \end{cases}
 \end{array}$$

Are their feasible regions convex?

The Hessian matrix I

$$(\text{Hess } f)_{ij} \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} \Rightarrow \text{Hess} \left(\frac{x^2}{y} \right) = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix}$$

Let f be a function of n variables and let $1 \leq i \leq n$ and $1 \leq j \leq n$. If the partial derivatives f'_i and f'_j of f exist in an open set S containing (x_1, \dots, x_n) and both of these partial derivatives are differentiable at (x_1, \dots, x_n) then $(\text{Hess } f)_{ij} = (\text{Hess } f)_{ji}$.

- If the Hessian at a given point has all positive eigenvalues, it is said to be a positive-definite matrix. This is the multivariable equivalent of “concave up” (we called it “convex” in this course).
- If all of the eigenvalues are negative, it is said to be a negative-definite matrix. This is like “concave down”. (we call it “concave” in this course).
- If the eigenvalues are mixed, you have a saddle point.

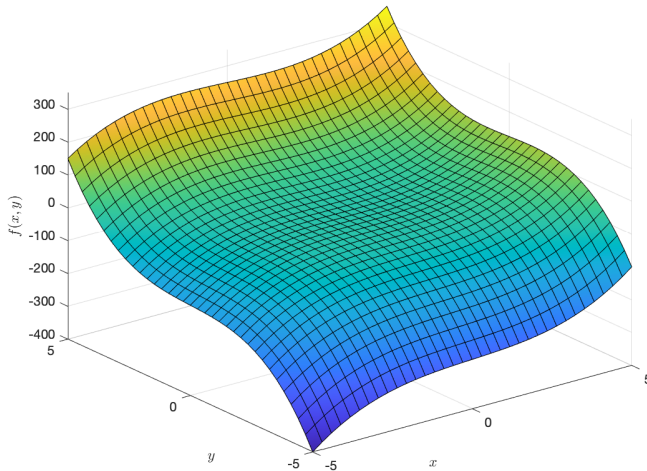
Exercise 5 Study the critical points of the function

$$f(x, y) = 4x + 2y - x^2 - 3y^2$$

Exercise 6 Check the character of the stationary point in $(0, 0)$ for the functions:

- $f(x, y) = x^2 + y^2$
- $g(x, y) = -x^2 - y^2$
- $h(x, y) = x^2 - y^2$
- $k(x, y) = x^3 + 2y^3 - xy$ (is it concave or convex in $(3, 3)$? what about in $(-3, -3)$?)

$$k(x, y) = x^3 + 2y^3 - xy$$



Unconstrained optimization of continuous $f : \mathbf{R} \rightarrow \mathbf{R}$

For unconstrained problems with just one variable x , if first and second derivatives exist in x^* :

Necessary conditions If $\frac{df}{dx} = 0$ at $x = x^*$

- $\frac{d^2f}{dx^2} \geq 0$, for a local minimum at $x = x^*$
- $\frac{d^2f}{dx^2} \leq 0$, for a local maximum at $x = x^*$

Sufficient conditions If $\frac{df}{dx} = 0$ at $x = x^*$

- $\frac{d^2f}{dx^2} > 0$, for a local minimum at $x = x^*$
- $\frac{d^2f}{dx^2} < 0$, for a local maximum at $x = x^*$

Unconstrained optimization of continuous $f : \mathbf{R}^n \rightarrow \mathbf{R}$

For unconstrained problems with several variables $\mathbf{x} = (x_1, \dots, x_n)$:

Sufficient condition If $\frac{\partial f}{\partial x_i} = 0$ at $\mathbf{x} = \mathbf{x}^*$:

- $\nabla f = \mathbf{0}$, for a local minimum at $\mathbf{x} = \mathbf{x}^*$ if the function is convex;
- $\nabla f = \mathbf{0}$, for a local maximum at $\mathbf{x} = \mathbf{x}^*$ if the function is concave.

Decision conditions The function f is a convex function if H_f is positive definite or positive semidefinite for all \mathbf{x} ; and f is concave if H_f is negative definite or negative semidefinite for all \mathbf{x} . For $f : \mathbf{R}^2 \rightarrow \mathbf{R}$

$$\mathbf{H}_f(x, y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} A & C \\ C & B \end{pmatrix}$$

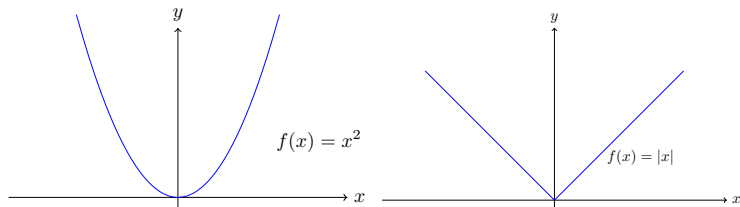
Unconstrained optimization of continuous $f : \mathbf{R}^n \rightarrow \mathbf{R}$

Let f be a twice-differentiable function of many variables on the convex open set S and denote the Hessian of f at the point x by $H(x)$. Then

- f is concave if and only if $H(x)$ is negative semidefinite $\forall x \in S$
- if $H(x)$ is negative definite $\forall x \in S$ then f is strictly concave (see Eq. 1)
- f is convex if and only if $H(x)$ is positive semidefinite $\forall x \in S$
- if $H(x)$ is positive definite $\forall x \in S$ then f is strictly convex (see Eq. 1)

Continuity vs derivability

Things are not so simple if, for example, functions are not derivable in x^* :



Or, if the function is not convex (nor concave), the sufficient condition is lost and we may find local optimal points, instead of global.

For constrained optimization problems, the shape of the feasible region adds difficulty.

Iterative methods: Summary

- 1 Nonlinear optimization
- 2 Concavity/convexity
- 3 Iterative methods**
- 4 Constrained optimization
 - Lagrange multipliers
 - Karush-Kuhn-Tucker conditions

One-Dimensional line search

Let us assume a concave function. First, establish error tolerance ϵ . Then:

- ① Set up lower and upper bounds for x by finding x_l, x_u such that $\frac{df}{dx_l} \geq 0$ and $\frac{df}{dx_u} \leq 0$
- ② Get new trial intermediate solution (eg, $x = \frac{x_l + x_u}{2}$)
- ③ If $x_u - x_l \leq \epsilon$, then terminate
- ④ If $\frac{df}{dx} \geq 0$, set $x_l = x$; if $\frac{df}{dx} \leq 0$, set $x_u = x$
- ⑤ Go to step 1

For an implementation, see [this link](#)

Golden search method I

For a continuous function it guarantees to find an extremum.

Golden ratio

Two quantities are in the golden ratio if their ratio is the same as the ratio of their sum to the larger of the two quantities:

$$\frac{a+b}{a} = \frac{a}{b} = \psi \Rightarrow \psi = \frac{1+\sqrt{5}}{2} = 1.616033988...$$

Procedure of the Golden search method, given a function to be minimized, $f(x)$, the interval to be searched as $x \in (x_1, x_4)$, and their functional values $f(x_1)$ and $f(x_2)$:

- 1 Calculate an interior point and its functional value $f(x_3)$. Ensure the position of x_3 follows the golden ratio between the distances

$$\frac{x_2 - x_1}{x_3 - x_1} = \frac{x_3 - x_1}{x_2 - x_3} = \psi.$$

Golden search method II

- 2 Using the triplet, determine if convergence criteria are fulfilled. If they are, estimate the x at the minimum from that triplet and terminate/return.
- 3 Select x_4 within the largest interval, following the same ratios as in Step 1.
- 4 The three points for the next iteration will be the one where $f(x)$ is a minimum, and the two points closest to it in x .
- 5 Go to step 2.

Exercise 7 Use the Golden Search method to find the minimum of function $f(x) = x^2 - 5x + 13$ in the interval $x \in (0, 10)$. Set up a spreadsheet to do so.

Exercises

Programming exercise 1

Build a Python code that can run one-dimensional search algorithms for user defined functions, using or not first derivatives. Implement the one-dimensional line search and the golden search algorithms and find the optimal solution for the functions:

- $f(x) = x^4 - 16x^3 + 45x^2 - 20x + 203$ within the range $x \in (2.5, 14)$
- $g(x) = x^5 - 2x^4 - 23x^3 - 12x^2 + 36x$ within the range $x \in (2, 3)$

Try both a direct implementation as well as the use of solvers in scipy.

Exercise 8 Draw the *Rosenbruck's function*,

$f(x, y) = 100(y - x^2)^2 + (1 - x)^2$ and discuss on its convexity, the feasibility of its minimization and the most appropriate numerical method to use.

Algorithms for unconstrained optimization

As a reference (not the central aim of this course) here is a list of some algorithms/methods to explore to perform unconstrained optimization in multiple dimensions:

- Newton's method,
- Quasi-Newton methods (e.g., BFGS),
- Trust-region methods,
- Conjugated gradient methods.

and their many variants.

Constrained optimization: Summary

- 1 Nonlinear optimization
- 2 Concavity/convexity
- 3 Iterative methods
- 4 Constrained optimization
 - Lagrange multipliers
 - Karush-Kuhn-Tucker conditions

In the previous section we have considered functions of several variables that are independent from each other. What happens if the variables are, however, not independent from each other? (see excellent material in [Richard Weber's web site](#))



What is the maximum height we can attain constrained to the trajectory of the road?

Lagrange multipliers I

Let us consider optimizing a function $f(x, y)$ (the height of the hill in the previous exemple) subject to the constrain $g(x, y) = c$ (the equation describing the path, the road).

One can simply substitute $g(x, y) = c$ in $f(x, y)$, but this may not involve trivial algebra.

Alternatively, let us consider the method of the Lagrange multipliers. To optimize $f(x, y)$ we require

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 0$$

we saw that if dx and dy were independent, $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$. But they are not here: they are constrained by $g(x, y)$, which is a constant:

$$dg = \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy = 0$$

Lagrange multipliers II

Let us consider a parameter λ (the *Lagrange multiplier*) and obtain an expression for which dx and dy are *indeed* independent:

$$d(f + \lambda g) = \left(\frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} \right) dx + \left(\frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} \right) dy = 0$$

Now:

$$\begin{aligned} \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} &= 0 \\ \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} &= 0 \\ g(x, y) &= c \end{aligned}$$

are sufficient to find λ , x^* and y^* .

The general problem we study takes the form

$$\begin{array}{ll} \text{maximize} & f(x) \\ \text{subject to} & g(x)=b \end{array}$$

where

$$\left. \begin{array}{ll} x \in \mathbf{R}^n & (n \text{ decision variables}) \\ f : \mathbf{R}^n \rightarrow \mathbf{R} & (\text{objective function}) \\ X \subseteq \mathbf{R}^n & (\text{regional constraints}) \\ g : \mathbf{R}^n \rightarrow \mathbf{R}^m \\ b \subseteq \mathbf{R}^m \end{array} \right\} (m \text{ functional constraints})$$

Regional and functional constraints define the **feasible set** for x

Case when $g(x) = 0$: The Lagrangian I

In the Lagrange approach, the constrained maximization (minimization) problem is rewritten as a Lagrange function

$$L(x, \lambda) = f(x) - \lambda g(x)$$

whose optimal point is a saddle point, i.e. a global maximum (minimum) over the domain of the choice variables and a global minimum (maximum) over the multipliers.

Let us take this example:

$$\begin{array}{ll} \text{maximize} & f(x) = x_1 - x_2 - 2x_3 \\ \text{subject to} & \begin{cases} x_1 + x_2 + x_3 = 5 \\ x_1^2 + x_2^2 = 4 \\ x = (x_1, x_2, x_3) \in \mathbf{R}^3 \end{cases} \end{array}$$

Case when $\mathbf{g}(\mathbf{x}) = 0$: The Lagrangian II

- ① We write the Lagrangian as (note we have two constraints):

$$\begin{aligned} L(\mathbf{x}, \lambda) &= x_1 - x_2 - 2x_3 - \lambda_1(x_1 + x_2 + x_3 - 5) - \lambda_2(x_1^2 + x_2^2 - 4) \\ &= [x_1(1 - \lambda_1) - \lambda_2 x_1^2] + [x_2(-1 - \lambda_1) - \lambda_2 x_2^2] + [x_3(-2 - \lambda_1) + 5\lambda_1 + 4\lambda_2] \end{aligned}$$

- ② We want to minimize the Lagrangian subject to the regional constraints $\mathbf{x} = (x_1, x_2, x_3) \in \mathbf{R}^3$.
- ③ We find the values λ such that will define a feasible set

$$Y = \left\{ \lambda : \underbrace{\min_{\mathbf{x} \in \mathbf{R}^3}}_{\substack{\text{min} \\ \mathbf{x} \in \mathbf{R}^3}} L(\mathbf{x}, \lambda) > -\infty \right\}$$

Case when $\mathbf{g}(\mathbf{x}) = 0$: The Lagrangian III

So, we only consider the values of λ for which we obtain a finite minimum. In the example

$$Y = \{\lambda : \lambda_1 = -2, \lambda_2 < 0\}$$

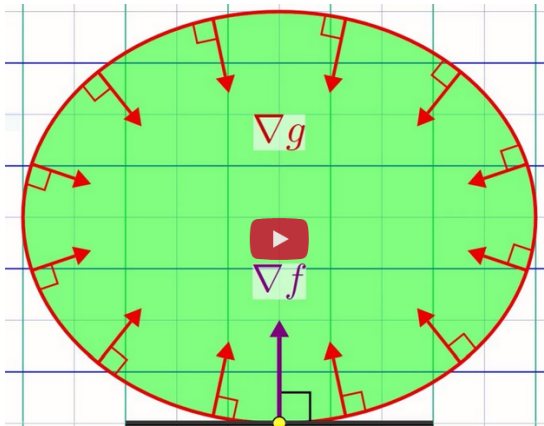
- 4 For $\lambda \in Y$, the minimum will be obtained at some $\mathbf{x}(\lambda)$ (that depends on λ in general). Here, $\mathbf{x}(\lambda) = (3/2\lambda_2, 1/2\lambda_2, x_3)^T$.
- 5 Finally, we find a λ value that provides a feasible value $\mathbf{x}(\lambda)$. Here

$$x_1^2 + x_2^2 = 4 \Rightarrow \lambda_2 = -\sqrt{5/8}$$

So, \mathbf{x}^* is optimal:

$$\mathbf{x}^* = (-3\sqrt{2/5}, -\sqrt{2/5}, 5 + 4\sqrt{2/5})^T; \lambda^* = (-2, -\sqrt{5/8})^T$$

Lagrange multipliers



Click on this image to access a video explaining the details of the Lagrange multipliers method.

Lagrange multipliers with two constraints (three variables) I

Consider optimizing $f(x, y, z)$ subject to two constraints $g_1(x, y, z) = c_1$ and $g_2(x, y, z) = c_2$.

The Lagrangian is:

$$\mathcal{L}(x, y, z, \lambda_1, \lambda_2) = f(x, y, z) - \lambda_1(g_1(x, y, z) - c_1) - \lambda_2(g_2(x, y, z) - c_2)$$

The necessary conditions for optimality are:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x} = 0 \\ \frac{\partial \mathcal{L}}{\partial y} = 0 \\ \frac{\partial \mathcal{L}}{\partial z} = 0 \\ g_1(x, y, z) = c_1 \\ g_2(x, y, z) = c_2 \end{cases}$$

Exercise 10 Use the Lagrange multipliers to solve this problem:

$$\begin{array}{ll}\text{minimize} & f(x, y) = xy + 1 \\ \text{subject to} & x^2 + y^2 = 1\end{array}$$

Exercise 11 Find the extrema of $f(x, y, z) = x^2 + y^2 + z^2$ subject to $x + y + z = 1$ and $x - y = 0$.

First Order Karush-Kuhn-Tucker Conditions I

The first order conditions for constrained optimization unify the treatment of equality and inequality constraints. They include:

- Allow for inequality constraints
- Allow for any number of constraints
- Constraints may be binding or not at the solution
- Allow for non-negative constraints (e.g., $x \geq 0$)
- Allow for boundary solutions (i.e., $x = 0$)
- Dual variables (Lagrange multipliers) and shadow values (marginal values) are included

Consider the problem: Maximize $f(x)$ subject to

$$x_1, x_2, \dots, x_n \geq 0$$

$$G^i(x) \leq b_i, \quad (i = 1, \dots, m)$$

First Order Karush-Kuhn-Tucker Conditions II

The Karush-Kuhn-Tucker (KKT) Conditions:

$\exists \lambda_1, \dots, \lambda_m \in \mathbb{R}_+$ such that (at \bar{x}) :

for $j = 1, \dots, n$:

$$\frac{\partial f}{\partial x_j} \leq \sum_{i=1}^m \lambda_i \frac{\partial G^i}{\partial x_j} \quad (\text{and } = \text{ if } \bar{x}_j > 0)$$

for $i = 1, \dots, m$:

$$G^i(\bar{x}) \leq b_i \quad (\text{and } = \text{ if } \lambda_i > 0)$$

First Order Karush-Kuhn-Tucker Conditions III

The Karush-Kuhn-Tucker (KKT) Conditions (vector form):

$\exists \lambda \in \mathbb{R}_+^m$ such that (at \bar{x}) :

$$\nabla f \leq \sum_{i=1}^m \lambda_i \nabla G^i$$

and

$$\bar{x} \cdot \left(\nabla f - \sum_{i=1}^m \lambda_i \nabla G^i \right) = 0$$

$$G(\bar{x}) \leq b$$

and

$$\lambda \cdot (G(\bar{x}) - b) = 0$$

These conditions generalize the method of Lagrange multipliers to include inequality constraints.

Exercise 12 Consider this NLO problem[1]:

$$\begin{array}{ll} \text{maximize} & f(x, y) = x^2y + 2y^2 \\ \text{subject to} & \begin{cases} x + 3y \leq 9 \\ x + 2y \leq 8 \\ 3x + 2y \leq 18 \\ 0 \leq x \leq 5 \\ 0 \leq y \leq 2 \end{cases} \end{array}$$

Solve the program graphically- Begin at point $(0, 0)$ and check the gradient. Check the KKT conditions. If they are not satisfied, find an improving direction in the feasible region.

The KKT conditions in practice I

In practice, if there exists a point $\mathbf{x}^* \in \Omega$ that solves the optimization problem, then \mathbf{x}^* satisfies the following Karush-Kuhn-Tucker (KKT) conditions:

1 Stationarity:

$$\lambda_0 \nabla f(\mathbf{x}^*) + \sum_{i=1}^p \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^q \mu_j \nabla h_j(\mathbf{x}^*) = 0$$

2 Primal feasibility:

$$g_i(\mathbf{x}^*) = 0 \quad \text{for } i = 1, \dots, p$$

$$h_j(\mathbf{x}^*) \leq 0 \quad \text{for } j = 1, \dots, q$$

The KKT conditions in practice II

3 Complementary slackness:

$$\mu_j h_j(\mathbf{x}^*) = 0 \quad \text{for } j = 1, \dots, q$$

Either the constraint is active ($h_j(\mathbf{x}^*) = 0$) or its multiplier is zero ($\mu_j = 0$).

4 Dual feasibility (sign condition):

$$\lambda_0 \geq 0, \quad \mu_j \geq 0 \quad \text{for all } j = 1, \dots, q \quad (\text{for minimization})$$

$$\mu_j \leq 0 \quad \text{for all } j = 1, \dots, q \quad (\text{for maximization})$$

Summary: The KKT conditions are necessary (and under convexity, sufficient) for optimality in constrained optimization. They combine stationarity, feasibility, complementary slackness, and dual feasibility.

Exercise 13 Consider the problem of maximizing $f(x, y, z) = x + y + z$ subject to the constraints:

$$h_1(x, y, z) = (y - 1)^2 + z^2 \leq 1, \quad h_2(x, y, z) = x^2 + (y - 1)^2 + z^2 \leq 3.$$

Use the Karush-Kuhn-Tucker (KKT) conditions to determine the optimal solution.

Exercise 14 Is it possible to maximize $f(x, y) = xy$ subject to $100 \geq x + y$ and $x \leq 40$?

Exercise 15 What about maximizing $f(x, y) = xy$ subject to $x + y^2 \leq 2$ and $x, y \geq 0$?

References

- [1] Michael W. Carter, Camille C. Price, and Ghaith Rabadi. Operations Research, 2nd Edition. CRC Press.
- [2] David Harel, with Yishai Feldman. Algorithmics: the spirit of computing, 3rd Edition. Addison-Wesley.
- [3] Ronald L. Rardin. Optimization in Operations Research, 2nd Edition. Pearson.
- [4] J. Hefferon. Linear algebra (4th Ed).
- [5] K.F. Riley, M.P. Hobson, S.J. Bence. Mathematical Methods for Physics and Engineering (2nd Ed). McGraw Hill.
- [6] J. Nocedal, S. J. Wright. Numerical Optimization (2nd Ed). Springer.
- [7] Kenneth J. Beers. Numerical methods for chemical engineering: applications in Matlab. Cambridge University Press.
- [8] D. Barber. Bayesian reasoning and machine learning. Cambridge University Press.