IE 426
Optimization models and applications

Lecture 19 — November 12, 2009

- Optimality conditions for NLP
- Lagrangian Relaxation
- Kuhn-Tucker conditions

Reading:

- Winston&Venkataramanan: Chapter 12, §§ 12.6, 12.8, 12.9
- Hillier&Lieberman: Chapter 14, §§ 14.4, 14.5[1], 14.6

[1] First paragraph only

Quadratic Programming (QP)

QPs are problems of the form

$$
\begin{array}{ll}
\min & x^\top Q^0 x + a_0 x \\
& x^\top Q^1 x + a_1 x \le b_1 \\
& x^\top Q^2 x + a_2 x \le b_2 \\
& \vdots \\
& x^\top Q^m x + a_m x \le b_m
\end{array}
$$

where $Q_0, Q_1 \ldots , Q_m$ are square matrices.

Quadratic Programming (QP)

QPs are problems of the form (equivalent)

$$
\begin{array}{ll}
\min & \sum_{i=1}^n \sum_{j=1}^n q^0_{ij} x_i x_j + \sum_{j=1}^n a_{0j} x_j \\
& \sum_{i=1}^n \sum_{j=1}^n q^1_{ij} x_i x_j + \sum_{j=1}^n a_{1j} x_j \le b_1 \\
& \sum_{i=1}^n \sum_{j=1}^n q^2_{ij} x_i x_j + \sum_{j=1}^n a_{2j} x_j \le b_2 \\
& \vdots \\
& \sum_{i=1}^n \sum_{j=1}^n q^m_{ij} x_i x_j + \sum_{j=1}^n a_{mj} x_j \le b_m
\end{array}
$$

where $Q_0, Q_1 \ldots , Q_m$ are square matrices.

When are they convex?

1. when the objective function is convex and
2. the constraints are convex.

All quadratic functions $\Rightarrow$ their Hessian is a constant matrix. The Hessian of the objective function is $2Q^0$, while the Hessian of the i-th constraint is $2Q^i$.

$\Rightarrow Q^0, Q^1, Q^2 \ldots , Q^m$ must be such that

- all their principal minors are nonnegative, or equivalently,
- they are positive semidefinite (PSD).

Examples

$$\begin{array}{llllllll}
\min & x^2 & +y^2 & -2xy & +2x & +5y & \\
& x^2 & & & +x & -3y & \le 4.2 \\
& x^2 & +y^2 & +2xy & +3x & & \le 10^4 \\
& x^2 & +y^2 & -4xy & & -y & \le -2124.1 \\
& x^2 & & -2xy & +x & +4y & \le 8 \cdot 10^7
\end{array}$$

Obj.f.: $f(x, y) = x^2 + y^2 - 2xy + 2x + 5y =$

$$= (x \;\; y) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + (2 \;\; 5) \begin{pmatrix} x \\ y \end{pmatrix},$$

therefore $Q^0 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ and $a_0 = (2 \;\; 5)$.

The Hessian of $x^2 + y^2 - 2xy + 2x + 5y$ is

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} = 2Q^0$$

Example (cont'd)

$Q^0 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Leading principal minors: 1,

$\begin{vmatrix} 1 & -1 \\ -1 & 1 \end{vmatrix} = 0.$

$Q^1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, ok

$Q^2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, ok

$Q^3 = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}$, $|Q^3| = -3 < 0$

$Q^4 = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}$, $|Q^4| = -1 < 0$

$\Rightarrow$ the problem is, in general, not convex.

Note: to check convexity in QP problems,

- we only need to check $Q^0 \ldots Q^m$ matrices
- we don't care about the linear terms $a_0 \ldots a_m$ or
- the right-hand sides of the constraints $b_1 \ldots b_m$

Examples

$$\begin{array}{lllllll}
\min & x^2 & +2y^2 & & & & \\
& -x^2 & -3y^2 & +xy & +x & -3y & \ge 24 \\
\text{[equivalent to} & x^2 & +3y^2 & -xy & -x & +3y & \le -24] \\
& x^2 & +y^2 & +2xy & +3x & & \le 125.56 \\
& & & & 3x & +y & = -104.23
\end{array}$$

$Q^0 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Leading principal minors: 1, 2.

$Q^1 = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 3 \end{pmatrix}$. Leading principal minors: 1, $\frac{11}{4}$.

$Q^2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Leading principal minors: 1, 0.

$\Rightarrow$ the problem is convex.

Examples

$$\begin{array}{llllllll}
\min & x_1^2 & +x_2^2 & +2x_3^2 & -x_1x_2 & +2x_1x_3 & +6x_2x_3 & \\
& x_1^2 & +2x_2^2 & +2x_3^2 & -4x_1x_2 & +x_1x_3 & -3x_2x_3 & \le 42 \\
& -x_1^2 & -x_2^2 & -x_3^2 & +4x_1x_2 & -8x_1x_3 & +2x_2x_3 & \ge -7 \\
\text{[equiv. to} & x_1^2 & +x_2^2 & +x_3^2 & -4x_1x_2 & +8x_1x_3 & -2x_2x_3 & \le 7
\end{array}$$

$Q^0 = \begin{pmatrix} 1 & -\frac{1}{2} & 1 \\ -\frac{1}{2} & 1 & 3 \\ 1 & 3 & 2 \end{pmatrix}$. Leading principal minors: 1, $\frac{3}{4}$,

$2 - \frac{3}{2} - \frac{3}{2} - 9 - \frac{2}{4} - 1 = -11.5$

$Q^1 = \begin{pmatrix} 1 & -2 & \frac{1}{2} \\ -2 & 2 & -\frac{3}{2} \\ \frac{1}{2} & -\frac{3}{2} & 2 \end{pmatrix}$. Leading principal minors: 1, -2, -3.75.

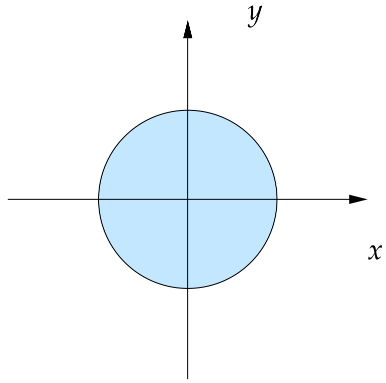
$Q^2 = \begin{pmatrix} 1 & -2 & 4 \\ -2 & 1 & -1 \\ 4 & -1 & 1 \end{pmatrix}$. Leading principal minors: 1, -3, -4.

$\Rightarrow$ the problem is, in general, nonconvex.

A little Geometry ...

$$x^2 + y^2 \leq 1$$

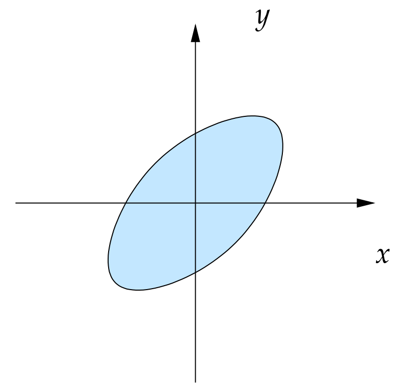
$Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. L.p.m: 1, 1 $\Rightarrow$ convex.



A little Geometry ...

$$5x^2 + 6y^2 + 6xy \leq 4$$

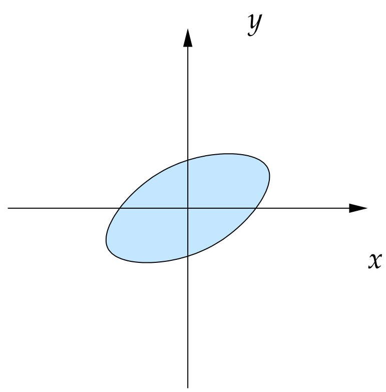
$Q = \begin{pmatrix} 5 & 3 \\ 3 & 6 \end{pmatrix}$. L.p.m: 30, 21 $\Rightarrow$ convex.



A little Geometry ...

$$17x^2 + 8y^2 + 12xy \leq 6$$

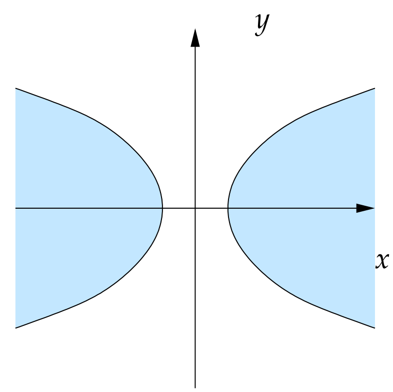
$Q = \begin{pmatrix} 17 & 6 \\ 6 & 8 \end{pmatrix}$. L.p.m: 136, 100 $\Rightarrow$ convex.



A little Geometry ...

$$x^2 - 4y^2 \leq 4$$

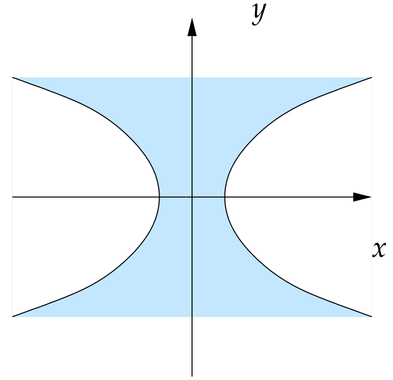
$Q = \begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix}$. L.p.m: 1, -4 $\Rightarrow$ non-convex.



A little Geometry . . .

$$x^2 - 4y^2 \geq 4, \quad or \quad -x^2 + 4y^2 \leq -4$$

$Q = \begin{pmatrix} -1 & 0 \\ 0 & 4 \end{pmatrix}$. L.p.m: -1, 4 $\Rightarrow$ non-convex.



One-variable Unconstrained Optimization

Consider a continuous and differentiable function $f(x_1)$ in one variable x_1, and the Optimization problem:

$$\min\ f(x_1)$$

No constraints, only one variable. A point $x_1^\star$ is a global optimum **only if**

$$\frac{df}{dx_1}(x_1^\star) = 0$$

The "only if" means "$\Rightarrow$", i.e.,

$$x_1^\star \text{ is a global optimum} \Rightarrow \frac{df}{dx_1}(x_1^\star) = 0.$$

That is, the global optimum has to be looked for in the set of points x_1 for which $\frac{df}{dx}(x_1) = 0$. If $\frac{df}{dx}(x_1^\star) = 0$ at a point $x_1^\star$, then not necessarily $x_1^\star$ is an optimum, not even a local one!

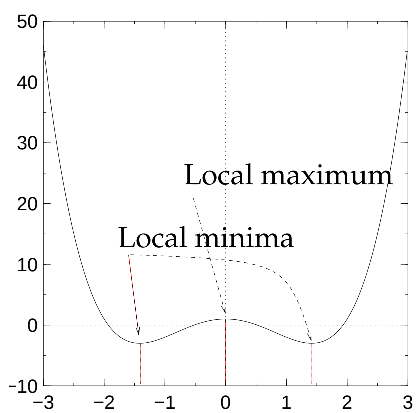
Example

$$\min f(x) = x^4 - 4x^2 + 1$$

What are the x such that

$$\frac{df}{dx} = 4x^3 - 8x = 0?$$

$$4x(x^2-2) = 0 \Leftrightarrow \begin{array}{ll} x = 0 & \vee \\ x = -\sqrt{2} & \vee \\ x = +\sqrt{2} & \end{array}$$



There are other conditions to check whether a point $x_1^\star$ such that which $\frac{df}{dx_1}(x_1^\star) = 0$ is a local minimum (e.g. second derivative), but let's not think about them now.

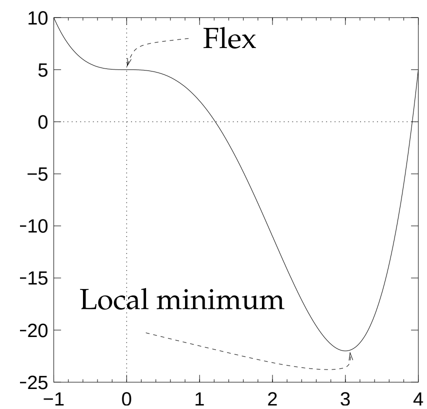
Example

$$\min f(x) = x^4 - 4x^3 + 1$$

What are the x such that

$$\frac{df}{dx} = 4x^3 - 12x^2 = 0?$$

$$4x^2(x-3) = 0 \Leftrightarrow \begin{array}{ll} x = 0 & \vee \\ x = 3 & \end{array}$$



(A flex is a point where the function changes from convex to concave or viceversa. In general, it is neither a local minimum nor a local maximum.)

Multi-variable Unconstrained Optimization

Consider a **continuous** and **differentiable** function $f(x_1, x_2 \dots, x_n)$ in n variables, and the Optimization problem:

$$\min \; f(x_1, x_2 \dots, x_n)$$

No constraints, n variables. A point $x = (x_1, x_2 \dots, x_n)$ is a global optimum **only if**

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0$$

The "only if" means "$\Rightarrow$", i.e.,

$$x \text{ is a global optimum} \Rightarrow \frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0.$$

That is, the global optimum has to be looked for in the set of points x for which $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0$.

Constrained (nonlinear) optimization

- The above conditions do **not** hold, in general, when we introduce constraints.
- Example: $\min\{x^2 + y^2 : x + y = 1, x \geq 0\}$ does not have a local minimum in $(0, 0)$: it's infeasible.

Consider

$$\begin{array}{lll} \min & f(x_1, x_2 \dots, x_n) & \\ & g_i(x_1, x_2 \dots, x_n) \geq a_i & i = 1, 2 \dots, p \\ & h_i(x_1, x_2 \dots, x_n) = b_i & i = p+1, p+2 \dots, m \end{array}$$

- We need to take into account those constraints
- We can relax all constraints, but in such a way that we still consider them somehow.

$\Rightarrow$ **Lagrangian Relaxation**

Relaxations

A relaxation of a minimization problem $\boxed{\min\{f(x) : x \in X\}}$ is a problem $\min\{f'(x) : x \in X'\}$ such that

- $X' \supseteq X$
- $f'(x) \leq f(x) \quad \forall x \in X$

A relaxation of a maximization problem $\boxed{\max\{f(x) : x \in X\}}$ is a problem $\max\{f'(x) : x \in X'\}$ such that

- $X' \supseteq X$
- $f'(x) \geq f(x) \quad \forall x \in X$

Lagrangian Relaxation

Consider a simple problem

$$(P) \qquad \begin{array}{ll} \min & f(x_1) \\ & h(x_1) = b \end{array}$$

and a **relaxation**:

$$(R) \qquad \min(f(x_1) + \lambda(h(x_1) - b))$$

Note: There is one optimization problem **for every value** of λ.

That is, we are optimizing over x_1 only, and consider λ an independent parameter. Is it a relaxation? Yes, because:

- if x_1 is feasible for (P), it is also feasible for (R); actually, any x_1 is feasible for (R) here
- for all x_1 feasible for (P), the objective function of R should be $\leq$ that of (P):

$$f(x_1) + \lambda(h(x_1) - b) = f(x_1) + \lambda \cdot (0) = f(x_1) \quad (\leq f(x_1))$$

Lagrangian Relaxation

Therefore, the quantity

$$\mathcal{L}(\lambda) = \min\{f(x_1) + \lambda(h(x_1) - b)\}$$

is a function of λ, not of x_1. For another simple problem

$$(P) \qquad \min \quad f(x_1) \qquad g(x_1) \geq a$$

we can again take a relaxation

$$(R) \qquad \min\{f(x_1) - \lambda(g(x_1) - a)\}$$

which is a relaxation for any **nonnegative** value of λ. Why?

- if x_1 is feasible for (P), it is also feasible for (R); again, any x_1 is feasible for (R) here
- for all x_1 feasible for (P), the objective function of R should be $\leq$ that of (P); remember λ is nonnegative

$$f(x_1) - \lambda(g(x_1) - a) = f(x_1) - \lambda \cdot (\geq 0) \leq f(x_1)$$

Lagrangian Relaxation

In order for $\mathcal{L}(\lambda)$ to be a relaxation, we look at the **sign** of the optimization (max or min) and of the constraints ($\leq$, $\geq$, or $=$).

- We actually relax the constraint(s), but penalize their violation with a multiplier λ.
- We can relax any subset of constraints, we just need to be careful with the signs of the λ's
- $\mathcal{L}(\lambda)$ gives a **lower bound** for any λ

Examples:

- $\min\{f(x) : g(x) \leq 0, h(x) = 0\}$ relaxes to $\min\{f(x) + \lambda g(x) : h(x) = 0\}$ for any $\lambda \geq 0$
- $\max\{f(x) : g(x) \leq 0, h(x) = 0, p(x) \leq 4\}$ relaxes to $\max\{f(x) + \lambda h(x) : g(x) \leq 0, p(x) \leq 4\}$ for any λ

Other examples

A Lagrangian Relaxation of	is	for any
$\min\{f(x) : g(x) = 0\}$	$\min\{f(x) + \lambda g(x)\}$	λ
$\min\{f(x) : g(x) = 0\}$	$\min\{f(x) - \lambda g(x)\}$	λ
$\min\{f(x) : g(x) \geq 0\}$	$\min\{f(x) + \lambda g(x)\}$	$\lambda \leq 0$
$\min\{f(x) : g(x) \geq 4\}$	$\min\{f(x) + \lambda(g(x) - 4)\}$	$\lambda \leq 0$
$\min\{f(x) : g(x) \geq 0\}$	$\min\{f(x) - \lambda g(x)\}$	$\lambda \geq 0$
$\min\{f(x) : g(x) \leq 1\}$	$\min\{f(x) + \lambda(g(x) - 1)\}$	$\lambda \geq 0$
$\max\{f(x) : g(x) = 0\}$	$\max\{f(x) + \lambda g(x)\}$	λ
$\max\{f(x) : g(x) = 0\}$	$\max\{f(x) - \lambda g(x)\}$	λ
$\max\{f(x) : g(x) \leq 0\}$	$\max\{f(x) - \lambda g(x)\}$	$\lambda \geq 0$
$\max\{f(x) : g(x) \leq 0\}$	$\max\{f(x) + \lambda g(x)\}$	$\lambda \leq 0$
$\max\{f(x) : g(x) \geq 0\}$	$\max\{f(x) + \lambda g(x)\}$	$\lambda \geq 0$

Back to NLP

Consider a problem with only equality constraints, for now:

$$\begin{array}{lll} (P) & \min & f(x_1, x_2 \ldots, x_n) \\ & & h_i(x_1, x_2 \ldots, x_n) = b_i \quad i = 1, 2 \ldots, k \end{array}$$

consider also the Lagrangian function

$$\begin{array}{rcl} \mathcal{L}(\lambda_1, \lambda_2 \ldots, \lambda_k) & = & f(x_1, x_2 \ldots, x_n) \\ & & +\lambda_1(h_1(x_1, x_2 \ldots, x_n) - b_1) \\ & & +\lambda_2(h_2(x_1, x_2 \ldots, x_n) - b_2) \\ & & \vdots \\ & & +\lambda_k(h_k(x_1, x_2 \ldots, x_n) - b_k) \end{array}$$
$$= f(x_1, x_2 \ldots, x_n) + \sum_{i=1}^{k} \lambda_i(h_i(x_1, x_2 \ldots, x_n) - b_i)$$

It gives a relaxation for any $(\lambda_1, \lambda_2 \ldots, \lambda_k) \in \mathbb{R}^k$.

Necessary conditions for optimality

$$\mathcal{L}(\lambda_1, \lambda_2 \dots, \lambda_k) = f(x_1, x_2 \dots, x_n) + \sum_{i=1}^{k} \lambda_i (h_i(x_1, x_2 \dots, x_n) - b_i)$$

Theorem: If a point $x^\star = (x_1^\star, x_2^\star \dots, x_n^\star)$ is optimal for (P), then

$$\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial \mathcal{L}}{\partial x_2} = \dots = \frac{\partial \mathcal{L}}{\partial x_n} = 0; \qquad \frac{\partial \mathcal{L}}{\partial \lambda_1} = \frac{\partial \mathcal{L}}{\partial \lambda_2} = \dots = \frac{\partial \mathcal{L}}{\partial \lambda_k} = 0.$$

The first part means $\frac{\partial \mathcal{L}}{\partial x_j} = 0 \ \forall j = 1, 2 \dots, n$, or

$$\frac{\partial f}{\partial x_j} + \sum_{i=1}^{k} \lambda_i \frac{\partial h_i}{\partial x_j} = 0 \qquad \forall j = 1, 2 \dots, n$$

The first row means $\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \ \forall i = 1, 2 \dots, k$, or

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \quad \Leftrightarrow \quad h_i(x_1, x_2 \dots, x_n) = b_i \quad \forall i = 1, 2 \dots, k$$

Necessary conditions for optimality

These are the Karush-Kuhn-Tucker (KKT) conditions.

If a point $x^\star = (x_1^\star, x_2^\star \dots, x_n^\star)$ is optimal for (P), then $\frac{\partial \mathcal{L}}{\partial x_i} = 0 \ \forall i = 1, 2 \dots, n$ and it is feasible.

$\Rightarrow$ global optimum found in points such that

$$\frac{\partial \mathcal{L}}{\partial x_i} = 0 \qquad \forall i = 1, 2 \dots, n$$

Ex. the tin can problem: $\min\{2\pi r^2 + 2\pi r h : \pi r^2 h = 20\}$

$$\mathcal{L}(\lambda) = 2\pi r^2 + 2\pi r h + \lambda(\pi r^2 h - 20)$$

$$\frac{\partial \mathcal{L}}{r} = 4\pi r + 2\pi h + 2\pi r h$$
$$\frac{\partial \mathcal{L}}{h} = 2\pi r + \lambda \pi r^2$$
$$\frac{\partial \mathcal{L}}{\lambda} = \pi r^2 h - 20$$

Back to NLP

Consider now a problem with inequality constraints:

$$\begin{array}{lll} (P) & \min & f(x_1, x_2 \dots, x_n) \\ & & g_i(x_1, x_2 \dots, x_n) \geq a_i \quad i = 1, 2 \dots, k \end{array}$$

consider also the Lagrangian function

$$\begin{aligned} \mathcal{L}(\lambda_1, \lambda_2 \dots, \lambda_k) \quad = \quad & f(x_1, x_2 \dots, x_n) \\ & -\lambda_1(g_1(x_1, x_2 \dots, x_n) - a_1) \\ & -\lambda_2(g_2(x_1, x_2 \dots, x_n) - a_2) \\ & \vdots \\ & -\lambda_k(g_k(x_1, x_2 \dots, x_n) - a_k) \\ = f(x_1, x_2 \dots, x_n) - \textstyle\sum_{i=1}^k & \lambda_i (g_i(x_1, x_2 \dots, x_n) - a_i) \end{aligned}$$

It gives a relaxation for any $(\lambda_1, \lambda_2 \dots, \lambda_k) \in \mathbb{R}_+^k$.

Necessary conditions for optimality

$$\mathcal{L}(\lambda_1, \lambda_2 \dots, \lambda_k) = f(x_1, x_2 \dots, x_n) - \sum_{i=1}^{k} \lambda_i (g_i(x_1, x_2 \dots, x_n) - a_i)$$

Theorem: If a point $x^\star = (x_1^\star, x_2^\star \dots, x_n^\star)$ is optimal for (P), then

$$\begin{aligned} & \frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial \mathcal{L}}{\partial x_2} = \dots = \frac{\partial \mathcal{L}}{\partial x_n} = 0; \\ & \lambda_1(g_1(x_1, x_2 \dots, x_n) - a_1) = \\ & \lambda_2(g_2(x_1, x_2 \dots, x_n) - a_2) = \\ & \vdots \\ & \lambda_k(g_k(x_1, x_2 \dots, x_n) - a_k) = 0; \\ & g_1(x_1, x_2 \dots, x_n) \geq a_1, \\ & g_2(x_1, x_2 \dots, x_n) \geq a_2, \\ & \vdots \\ & g_k(x_1, x_2 \dots, x_n) \geq a_k; \\ & \lambda_1, \lambda_2 \dots, \lambda_k \geq 0. \end{aligned}$$

Necessary conditions for optimality

KKT optimality conditions in the general case:

$$
\begin{array}{lll}
(1) & \frac{\partial \mathcal{L}}{\partial x_i} = 0 & \forall i = 1, 2 \ldots n \\
(2) & \lambda_i (g_i(x_1, x_2 \ldots, x_n) - a_i) = 0 & \forall i = 1, 2 \ldots k \\
(3) & g_i(x_1, x_2 \ldots, x_n) \geq a_i & \forall i = 1, 2 \ldots k \\
(4) & \lambda_i \geq 0 & \forall i = 1, 2 \ldots k
\end{array}
$$

(1) means $\frac{\partial \mathcal{L}}{\partial x_i} = 0 \ \forall i = 1, 2 \ldots, n$, or

$$\frac{\partial f}{\partial x_i} - \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i} = 0 \qquad \forall i = 1, 2 \ldots, n$$

(2) means that, for all $i = 1, 2 \ldots, k$, that either $\lambda_i = 0$ or the constraint is satisfied at equality.

(3) means that all constraints should be satisfied (obvious)

(4) means that we have a relaxation

Convex case

If $f(x_1, x_2 \ldots, x_n)$ convex, $-g_i(x_1, x_2 \ldots, x_n)$ are convex, and $h_1(x_1, x_2 \ldots, x_n)$ are affine (linear), then any point satisfying the KKT conditions is a global optimum.

Example:

$$
\begin{array}{ll}
\min & x^2 + 2y^2 \\
 & x + y \geq 3 \\
 & y - x^2 \geq 1
\end{array}
$$

Caveat:

- the books may use different notation
- there are further conditions to ensure optimality, called *constraint qualification*; we won't see them