LECTURE 6: CONSTRAINED OPTIMIZATION – OPTIMALITY CONDITIONS

1. Basic concepts
2. Necessary conditions – KKT conditions
3. Sufficient conditions

Constrained optimization

- General form:

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{s.t.} & \\
& \left.\begin{array}{l} h_1(x) = 0 \\ \vdots \\ h_m(x) = 0 \\ g_1(x) \le 0 \\ \vdots \\ g_p(x) \le 0 \end{array}\right\} \begin{array}{l}\text{functional constraints} \\ \text{(explicit)}\end{array} \\
& x \in \Omega \subset E^n, \ \text{(set/implicit constraints)}
\end{array}
$$

where $m \le n$ and $f, \ h_i, \ g_j \in C^k$ (most likely $k = 1$ or 2).

Short form:

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{s.t} & h(x) = 0 \\
& g(x) \le 0 \\
& x \in \Omega,
\end{array}
$$

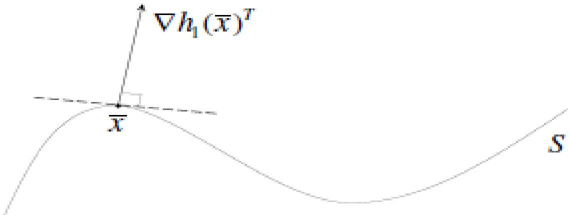
where $h(x) = (h_1(x), \cdots, h_m(x))^T$ and $g(x) = (g_1(x), \cdots, g_p(x))^T$.

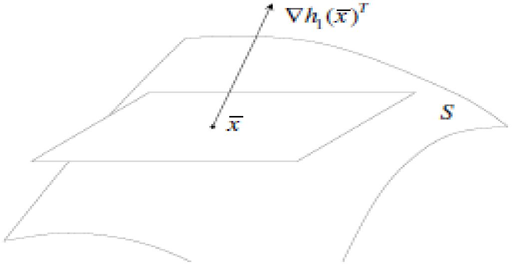
Basic concepts

- Definition:

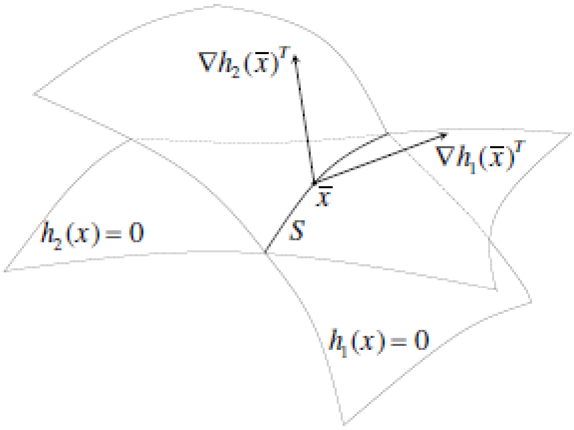
Let $h(\cdot) = (h_1(\cdot), \cdots, h_m(\cdot))^T$ with $h_i : E^n \to R$. Then $S \triangleq \{x \in E^n \mid h(x) = 0\}$ is a <u>surface</u>. When each $h_i(\cdot)$ is a (C^k) smooth function, then S is a <u>(C^k) smooth surface</u>.

- For $m = 1, h = h_1$,





- For $m = 2, h = \begin{pmatrix} h1 \\ h2 \end{pmatrix}$,



- **<u>Observations:</u>**

1. Given $\bar{x} \in S$, $\nabla h_i(\bar{x})$ is orthogonal to a "tangent" plane of S at $\bar{x}$ for each i .

2. The feasible directions at $\bar{x}$ falls in

   $T(\bar{x}) \triangleq \{d \in E^n \mid \nabla h_i(\bar{x})d = 0,\ i = 1, 2, \cdots, m\}$.

   We may write

   "$\nabla h(\bar{x})d = 0$" for "$\nabla h_i(\bar{x})d = 0,\ i = 1, \cdots, m$".

Basic concepts

- Definition

Let $\bar{x} \in S \triangleq \{x \in E^n \mid h(x) = 0\}$. Then $\bar{x}$ is a <u>regular point</u> if the gradient vectors $\{\nabla h_1(\bar{x}), \cdots, \nabla h_m(\bar{x})\}$ are linearly independent.

- **Observations:**

1. Every point $x \in E^n$ s.t. $h(x) = 0$ is "relatively interior" to $S = \{x \in E^n \mid h(x) = 0\}$.

2. When $\bar{x}$ is a regular point, $T(\bar{x}) \triangleq \{d \in E^n \mid \nabla h(\bar{x}) d = 0\}$ is the tangent plane at $\bar{x}$ with explicit geometric meanings.

Basic concepts

- Definition

(i) A curve $\mathscr{C}$ on a surface S is a set of points $x(t) \in S$ continuously parameterized by t over an interval $[a, b]$, i.e.,
$\mathscr{C} = \{\ x(t) \in S \mid t \in [a, b]\ \}$.

(ii) $\mathscr{C}$ is differentiable, if $\dot{x} \triangleq \frac{dx(t)}{dt}$ exists.

(iii) $\mathscr{C}$ is twice differentiable, if
$\ddot{x} \triangleq \frac{d}{dt}(\frac{dx(t)}{dt})$ exists.

(iv) $\mathscr{C}$ passes through $\bar{x} \in S$, if $\exists\ \bar{t} \in (a, b)$, s.t. $x(\bar{t}) = \bar{x}$.

In this case, $\dot{x}(\bar{t})$ is the derivative of $\mathscr{C}$ at $\bar{x}$.

(v) The tangent plane at $\bar{x} \in S$ is the collection of the derivatives at $\bar{x}$ of all differentiable curves passing through $\bar{x}$.

Basic concepts

- Theorem:

Let $\bar{x}$ be a regular point of the surface $S \triangleq \{x \in E^n \mid h(x) = 0\}$. Then the tangent plane is equal to

$$T(\bar{x}) \triangleq \{d \in E^n \mid \nabla h(\bar{x}) d = 0\}.$$

- Proof:

Luenberger P. 298.

First order necessary conditions

- NLP with equality constraints
- Theorem:

Let x^* be a regular point of $S = \{x \in E^n \mid h(x) = 0\}$ and a local minimum (maximum) point of f over S. Then

$$\nabla f(x^*)d = 0$$

for all $d \in T(x^*) = \{d \in E^n \mid \nabla h(x^*)d = 0\}$.

- Proof:

Directly from Taylor's Theorem, or Luenberger P. 300.

First order necessary conditions

- Corollary: Let x^* be a regular point of $S = \{x \in E^n \mid h(x) = 0\}$ and a local minimum (maximum) point of f over S. Then $\exists\ \lambda \in E^m$ such that

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0.$$

- Proof:

Consider the following LP problem

$$\begin{array}{ll} \text{minimize} & -\nabla f(x^*)d \\ \text{s.t.} & \nabla h(x^*)d = 0 \\ & d \in E^n \end{array}$$

and its dual problem

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{s.t.} & \nabla h(x^*)^T \lambda = -\nabla f(x^*)^T \\ & \lambda \in E^m. \end{array}$$

Since x^* is a regular point, the previous theorem implies that the dual problem is feasible. Hence $\exists\ \lambda \in E^m$, s.t.

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0.$$

Observations

1. We may define

$$\ell(x,\lambda) \triangleq f(x) + \lambda^T h(x)$$

as the Lagrangian associated with the constrained optimization problem. And we may call λ the Lagrange/Lagrangian vector and λ_i the Lagrange/Lagrangian multiplier associated with $h_i(x) = 0$.

2. The necessary conditions can be expressed as

$$\nabla_x \ell(x,\lambda) \left(= \nabla f(x) + \lambda^T \nabla h(x)\right) = 0,$$

$$\nabla_\lambda \ell(x,\lambda) \left(= h(x)\right) = 0,$$

which is a system of $n+m$ variables satisfying $n+m$ equations.

First order necessary conditions

- NLP with equality and inequality constraints

$$
\text{(NLP)} \qquad \begin{array}{ll} \text{maximize} & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \leq 0 \\ & x \in E^n \end{array}
$$

- Definition

Let $\bar{x}$ be a feasible solution and $J(\bar{x})$ be the index set of all active (inequality) constraints.

$\bar{x}$ is said to be a regular point if the gradient vectors $\nabla h_i(\bar{x})$, $i = 1, 2, \cdots, m$, and $\nabla g_j(\bar{x})$, $j \in J(\bar{x})$, are linearly independent.

Main theorem

- Theorem (KKT Conditions)

Let x^* be a relative minimum point for (NLP) that is a regular point. Then $\exists$ a vector $\lambda \in E^m$ and a vector $\mu \in E^p_+$ s.t.

$$(*) \begin{cases} \nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0 \\ \mu^T g(x^*) = 0. \end{cases}$$

Proof

Since x^* is known, we know each inequality constraint is active or inactive. From the first-order necessary conditions for equality constraints, $\exists\ \lambda^T \in E^m$ and $\mu_j \in R$, $j \in J(x^*)$, such that

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \sum_{j \in J(x^*)} \mu_j \nabla g_j(x^*) = 0.$$

Taking $\mu_j = 0$ for $j \notin J(x^*)$ leads to $(*)$ except that $\mu_j \in R$ for $j \in J(x^*)$.

Let $J^+ = \{j \in J(x^*),\ \mu_j \geq 0\}$ and

$J^- = \{j \in J(x^*),\ \mu_j < 0\}$.

If $J^- \neq \phi$, then we have

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \sum_{j \in J^+} \mu_j \nabla g_j(x^*) = \sum_{j \in J^-} (-\mu_j) \nabla g_j(x^*).$$

W.L.O.G. we may assume that $\nabla g_j(x^*) \neq 0$, $\forall\ j \in J^-$. Remember x^* is a regular point. A negative direction of the right hand side vector projected onto the null space of $\{\nabla h_i(x^*)\}$ leads to a direction d that has components in $-\nabla g_j(x^*)$ $(j \in J)$ and $-\nabla f(x^*)$ only.

In this case, moving along d from x^* will

(1) reduce the value in f,

(2) reduce the value in $g_j, j \in J$,

(3) remain the same value in h_i, $i = 1, \cdots, m$.

This contradicts to the fact that x^* is a local minimum point. Hence $J^- = \phi$ and $\mu_j \geq 0$, $\forall\ j \in J$.

Common terminologies

1.

$$\left.\begin{aligned} h(x^*) &= 0 \\ g(x^*) &\le 0 \end{aligned}\right\}\text{primal feasibility (PF)}$$

$$\left.\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) &= 0 \\ \lambda_i \in R, \quad \mu_j \ge 0 \end{aligned}\right\}\text{dual feasibility } (DF)$$

$\mu^T g(x^*) = 0\}$ complementary slackness (CS).

(PF) + (DF) + (CS) = K-K-T conditions.

2. Any point $\bar{x} \in E^n$ for which $\exists\ (\lambda, \mu)$ s.t. $(\bar{x}, \lambda, \mu)$ satisfies K-K-T conditions is called a K-K-T point.

3. The requirement of " x^* is a regular point" is also called "Independence Constraint Qualification" (ICQ). There are many kinds of constraint qualification (CQ) conditions that relaxes (ICQ).

Constraint qualifications (CQ)

- Slater's condition (Slater's CQ)

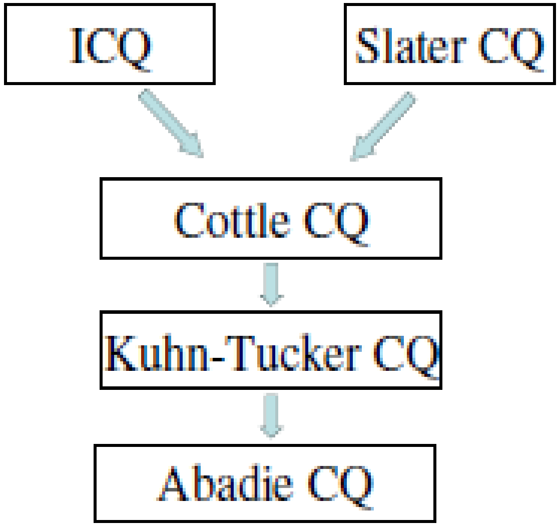
(i) Each g_j is continuous and pseudo convex at $\bar{x}$, $j \in J(\bar{x})$.

(ii) Each h_i is quasi convex, quasi concave, and continuous differentiable at $\bar{x}$.

(iii) $\{\nabla h_i(\bar{x})\}$ are linearly independent.

(iv) $\exists\, x$ s.t. $g_j(x) < 0, \forall\, j \in J(\bar{x})$, and $h_i(x) = 0, \forall\, i$.

Other CQs



Second order necessary conditions

- Observations:

1. Following the corollary, at x^* there exists a $\lambda \in E^m$, s.t. $\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$. Notice that x^* is regular and $T(x^*) = \{y \in E^n \mid \nabla h(x^*)y = 0\}$ is a subspace of E^n.

   Consider the problem of minimizing $\tilde{f}(y) \triangleq f(x^* + y) + \lambda^T h(x^* + y)$ over $T(x^*)$.

   We know for sure that 0 is a local minimizer for this unconstrained problem in $T(x^*)$ space if and only if x^* is a local minimizer of $f(x)$ over $S = \{x \in E^n \mid h(x) = 0\}$.

2. The second-order necessary conditions for unconstrained optimization problems require that

   $$\tilde{F}(0) = F(x^* + 0) + \lambda^T H(x^* + 0) \triangleq L(x^*)$$

   be positive semidefinite on $T(x^*)$, i.e.,

   $$y^T L(x^*) y \geq 0, \quad \forall y \in T(x^*).$$

Result 1

Theorem (2nd Order Necessary Conditions / Equality Constraints)

Let x^* be a local minimum point of f over $S = \{x \in E^n \mid h(x) = 0\}$ and x^* is a regular point. Then $\exists$ a $\lambda \in E^m$ s.t.

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

and the matrix

$$L(x^*) = F(x^*) + \lambda^T H(x^*)$$

is positive semidefinite on

$$T(x^*) \triangleq \{y \mid \nabla h(x^*) y = 0\}.$$

Proof: (Luenberger P. 306-307 has an equivalent derivation.)

Result 2

- **Theorem** (2nd-order Necessary Conditions / Equality and Inequality Conditions)

  Let $f, g, h \in C^2$ and x^* be a regular point of $\mathscr{F} = \{x \in E^n \mid h(x) = 0,\ g(x) \leq 0\}$. If x^* is a local minimizer of f over $\mathscr{F}$, then $\exists\ \lambda \in E^m,\ \mu \in E^p_+$, such that

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0,$$
$$\mu^T g(x^*) = 0.$$

  and

$$L(x^*) = F(x^*) + \lambda^T H(x^*) + \mu^T G(x^*)$$

  is positive semidefinite on the tangent subspace of all the active constraints at x^*

  Proof: Direct consequence of the same logic used in the previous theorem.

Sufficient conditions for optimality

- Key idea:

  Following the 2nd-order sufficient conditions

  for unconstrained optimization problem will lead to an

  answer to the constrained case.

Result 1

- **Theorem:** (2nd-order Sufficient Conditions / Equality Constraints)

  Let $x^* \in E^n$ and $\lambda \in E^m$ s.t.

  $$\begin{cases} h(x^*) = 0 \\ \nabla f(x^*) + \lambda^T h(x^*) = 0. \end{cases}$$

  If $L(x^*) = F(x^*) + \lambda^T H(x^*)$ is positive definite on $T(x^*) = \{y \in E^n \mid \nabla h(x^*)y = 0\}$, then x^* is a strict local minimum point of f over

  $$S = \{x \in E^n \mid h(x) = 0\}.$$

Proof:

Luenberger P. 307 proved explicitly by contradiction.

Observations

1. When inequality constraints involved, the index set $J(x^*) = \{j \mid g_j(x^*) = 0\}$ and

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$$

with $\lambda \in E^m$ and $\mu_j \geq 0, \ \forall j \in J(x^*)$.

If $\mu_j = 0$, then g_j actually plays no role for an active constraint. In this case, we call it a "degenerate inequality " to begin with. We define

$$\bar{J}(x^*) = \{j \mid g_j(x^*) = 0 \ \text{ and } \mu_j > 0\}$$

to index those "nondegenerate" inequalities with positive Lagrange multipliers.

2. Note that $J(x^*) = \bar{J}(x^*)$ under the "nondegeneracy assumption".

3. The 2nd-order sufficient conditions work on $\bar{J}(x^*)$ to avoid degeneracy.

Result 2

Theorem: (2nd-order Sufficient Conditions /Equality and Inequality Conditions)

Let $f, g, h \in C^2$ and $x^* \in \mathscr{F}$. If $\exists\ \lambda \in E^m$, $\mu \in E^p_+$, s.t.

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0,$$
$$\mu^T g(x^*) = 0,$$

and the Hessian matrix

$$L(x^*) = F(x^*) + \lambda^T H(x^*) + \mu^T G(x^*)$$

is positive definite on the subspace

$$\bar{T}(x^*) = \{y \in E^n \mid \nabla h(x^*) y = 0,$$
$$\nabla g_j(x^*) y = 0,\ \forall j \in \bar{J}(x^*)\},$$

then x^* is a strict relative minimizer of f over $\mathscr{F}$.

Observations

1. $T(x^*) \subset \bar{T}(x^*)$
2. $T(x^*) = \bar{T}(x^*)$ iff every active inequality constraint at x^* is nondegenerate.

Corollary:

Let f be strictly convex, g_j convex, and x^* a K-K-T point of minimizing $f(x)$ over $\{x \in E^n \mid g(x) \leq 0\}$. Then x^* is a global minimizer.

Interesting questions

Consider the following two problems:

$$(P_1) \quad \begin{array}{ll} \text{Min} & f(x) \\ \text{s. t.} & g_1(x) \le 0 \\ & \vdots \\ & g_k(x) \le 0 \\ & x \in E^n \end{array}$$

$$(P_2) \quad \begin{array}{ll} \text{Min} & \bar{f}(x,s) = f(x) \\ \text{s. t.} & h_1(x,s) \triangleq g_1(x) + s_1^2 = 0 \\ & \vdots \\ & h_k(x,s) \triangleq g_k(x) + s_k^2 = 0 \\ & (x,s) \in E^{n+k} \end{array}$$

It is clear that (P_1) and (P_2) are equivalent, but (P_1) has k inequality constraints and (P_2) has k equality constraints.

Question 1: Why should we explicitly use the (KKT) conditions of (P_1)?

Question2:

Can the (KKT) conditions of (P_1) be derived for the (KKT) conditions of (P_2)?

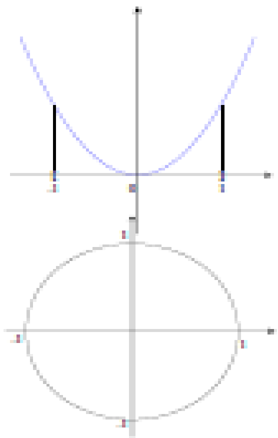
Answer to question 1

Question 1: Why should we explicitly use the (KKT) conditions of (P_1)?

Example:

$$(P_1) \quad \begin{array}{ll} \text{Min} & x^2 \\ \text{s. t.} & x^2 - 1 \leq 0 \end{array}$$

$$(P_2) \quad \begin{array}{ll} \text{Min} & x^2 + 0 \cdot s^2 \\ \text{s. t.} & x^2 - 1 + s^2 = 0 \end{array}$$



Necessary Conditions

$$(P_1) \quad \begin{cases} 2x + \mu(2x) = 0 \quad , \quad \mu \geq 0 \\ \mu(x^2 - 1) = 0 \end{cases}$$

$$2x(1 + \mu) = 0, \ \mu \geq 0$$

$$\Rightarrow \ x = 0 \text{ and } \mu = 0.$$

$$(P_2) \quad \begin{cases} \begin{bmatrix} 2x \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 2x \\ 2s \end{bmatrix} = 0 \quad , \quad \lambda \in R \\ x^2 + s^2 = 1 \end{cases}$$

$$\begin{cases} 2x(1+\lambda) & = 0 \\ 2\lambda s & = 0 \\ x^2 + s^2 & = 1 \end{cases} \Rightarrow \underbrace{\begin{array}{c} \lambda = 0 \\ x = 0 \\ s = \pm 1 \end{array}}_{\text{real solutions}} \quad \text{or} \quad \underbrace{\begin{array}{c} s = 0 \\ x = \pm 1 \\ \lambda = -1 \end{array}}_{\text{false solutions}}$$

Answer to question 1

<u>Sufficient Conditions</u>

(P_1)

$F(x) = (2)$, $G(x) = (2)$.

At $x = 0$ and $\mu = 0$,

$$F(x) + \mu G(x) = (2) > 0 \quad \text{(p.d.)}$$

(P_1) is a convex programming problem, (KKT) conditions are sufficient for optimality.

(P_2)

$F(x, s) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, $H(x, s) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

At $x = 0$, $s = \pm 1$ and $\lambda = 0$

$$F(x, s) + \lambda H(x, s) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{(p.s.d.)}$$

At $x = \pm 1$, $s = 0$ and $\lambda = -1$

$F(x, s) + \lambda H(x, s)$

$$= \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \quad \text{(n.s.d.)}$$

(P_2) is not a convex programming problem, (KKT) conditions are not sufficient for optimality.

Answer to question 2

Question2:

Can the (KKT) conditions of (P_1) be derived for the (KKT) conditions of (P_2)?

Answer:

(KKT) conditions of (P_1):

$$\nabla f(x) + \sum_{j=1}^{k} \mu_j \nabla g_j(x) = 0 \quad , \ \mu_j \geq 0$$
$$\mu_j g_j(x) = 0$$

(KKT) conditions of (P_2):

$$\nabla \tilde{f}(x,s) + \sum_{j=1}^{k} \lambda_j \nabla h_j(x,s) = 0 \quad , \ \lambda_j \in R$$
$$h_j(x,s) = 0$$

Notice that for problem (P_2),

$$\nabla \tilde{f}(x,s) = (\ \nabla f(x), 0\ )$$
$$\nabla h_j(x,s) = (\ \nabla g_j(x), 0, \cdots, 2s_j, 0, \cdots, 0\ )$$
$$\Rightarrow \lambda_j(2s_j) = 0, \quad j = 1, 2, \cdots, k.$$

- If $\lambda_j > 0$, then $s_j = 0$ and, consequently, $h_j(x,s) = g_j(x) = 0$ and $\lambda_j g_j(x) = 0$.
- If $\lambda_j = 0$, then $\lambda_j g_j(x) = 0$ is obvious.

W.L.O.G. we say $\lambda_1 < 0$ and $\lambda_2, \cdots, \lambda_k \geq 0$ and x is a strick local minimum solution.

Then

$$\begin{pmatrix} \nabla f(x)^T \\ 0 \end{pmatrix} + \sum_{j=2}^{k} \lambda_j \begin{pmatrix} \nabla g_j(x) \\ 0 \\ \vdots \\ 2s_j \\ \vdots \\ 0 \end{pmatrix} = -\lambda_1 \begin{pmatrix} \nabla g_1(x) \\ 2s_1 = 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In other words,

$$-\nabla f(x) + \sum_{j=2}^{k} \overset{\geq 0}{\lambda_j}(-\nabla g_j(x)) = -(\overset{> 0}{-\lambda_1})\nabla g_1(x)$$

Answer to question 2

Note that $g_j(x) \leq 0$, moving along the direction of $-\nabla g_1(x)$ will reduce $g_1(x)$, reduce or retain $g_j(x)$, $j = 2, \cdots, k$, and reduce $f(x)$. This contradicts the assumption that x is a local minimizer. Hence $\lambda_j \geq 0$ for $j = 1, 2, \cdots, k$. Moreover, we know that $\lambda_j g_j(x) = 0$. Therefore, we can choose $\mu_j = \lambda_j$, for $j = 1, 2, \cdots, k$ to satisfy the (KKT) conditions for (P_1).