5 Nonlinear Programming

General form:

$$\text{(N) Minimize (Maximize) } f(x) \text{ subject to } h(x) = 0 \text{ and } g(x) \leq 0,$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h = (h_1, \ldots, h_m) : \mathbb{R}^n \to \mathbb{R}^m$, $g = (g_1, \ldots, g_p) : \mathbb{R}^n \to \mathbb{R}^p$.
Notions associated to (N)

- *Feasible set*: $S = \{x \in \mathbb{R}^n \mid h(x) = 0 \text{ and } g(x) \leq 0\}$

(Min)
- $x^* \in \mathbb{R}^n$ *(strict) optimal solution* if $\forall x \in S$ $(x \neq x^*)$ $f(x) \underset{(>)}{\geq} f(x^*)$

(Max)
- $x^* \in \mathbb{R}^n$ *local (strict) optimal solution* if $\exists \varepsilon_0 > 0, \forall x \in S, \|x - x^*\| < \varepsilon_0, (x \neq x^*)$ $f(x^*) \underset{(>)}{\geq} f(x)$

5.1 Lagrange Multipliers (Sufficient Conditions)

$$\text{(NE) Minimize } f(x) \text{ subject to } h(x) = 0,$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h = (h_1, \ldots, h_m) : \mathbb{R}^n \to \mathbb{R}^m$, and $m \leq n$.

Notions associated to (NE)

- A feasible point x^* (i.e., $h(x^*) = 0$) is a *regular point* of the constraints if

$\nabla h_1(x^*), \nabla h_2(x^*), \ldots, \nabla h_m(x^*)$ are linearly-independent

In terms of the *Jacobian matrix* $Dh(x^*) = (\frac{\partial h_i}{\partial x_j})_{\substack{i=\overline{1,m} \\ j=\overline{1,n}}} = \begin{bmatrix} Dh_1(x^*) \\ \vdots \\ Dh_m(x^*) \end{bmatrix} = \begin{bmatrix} \nabla h_1(x^*)^T \\ \vdots \\ \nabla h_m(x^*)^T \end{bmatrix}$

x^* is regular iff rank $Dh(x^*) = m$ - full rank (dim row space)

Example. Identify h and find the regular/non-regular points for the equations

$h : \mathbb{R}^3 \to \mathbb{R}^2, h = (h_1, h_2)$ $\quad x_1x_2 + x_2x_3 + x_3x_1 = a, \quad x_1 + x_2 + x_3 = b.$

$h_1(x_1, x_2, x_3) = x_1x_2 + x_2x_3 + x_3x_1 - a$, $h_2(x_1, x_2, x_3) = x_1 + x_2 + x_3 - b$

$\nabla h_1(x) = \begin{bmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{bmatrix}, \nabla h_2(x) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Q. When are 2 vectors linearly dep?
A. When they are proportional!

$\{\nabla h_1(x), \nabla h_2(x)\}$ lin dep iff $x_2 + x_3 = x_1 + x_3 = x_1 + x_2$ iff $x_1 = x_2 = x_3$

Non-Reg Points $:= \{x = \overset{(\alpha,\alpha,\alpha)}{(x_1, x_2, x_3)} \mid x_1 = x_2 = x_3 = \alpha \in \mathbb{R}\}$

Reg Points $:= \{x = (x_1, x_2, x_3) \mid x_1 \neq x_2 \text{ or } x_2 \neq x_3 \text{ or } x_3 \neq x_1\}$

- The *tangent space* at x^* on the surface $S := \{x \mid h(x) = 0\}$

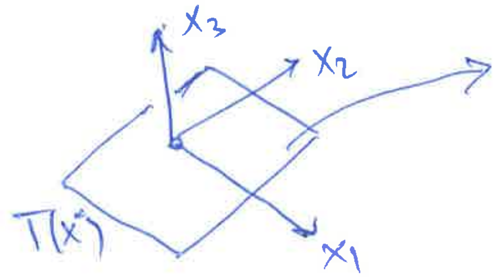
$T(x^*) = \{y \in \mathbb{R}^n \mid Dh(x^*)y = 0\}$ = Null Space of $Dh(x^*) = \mathcal{N}(Dh(x^*))$

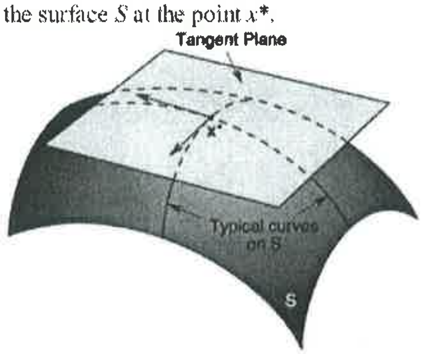
Note that $0 \in T(x^*)$ and when x^* is regular $\dim T(x^*) = n - m$ bc $\underbrace{\dim \text{Row}\, Dh(x^*)}_{\text{rank}\, Dh(x^*) = m} + \dim \mathcal{N}(Dh(x^*)) = n$

The *tangent plane* at x^*

$$TP(x^*) = T(x^*) + x^* = \{x^* + y \mid y \in T(x^*)\}$$

Tangent plane to the surface S at the point x^*.





Theorem. Suppose that $x^* \in S = \{x \mid h(x) = 0\}$ is a regular point and $T(x^*)$ is the tangent space at x^*. Then, $y \in T(x^*)$ if and only if there exists a differentiable curve in S passing through x^* with derivative y at x^*, (that is, $\exists x : I \to \mathbb{R}^n$, $x(I) \subset S$, $\exists t^* \in I :\ x(t^*) = x^*$, $\frac{dx}{dt}(t^*) = y$).

Proof. ($\Leftarrow$) Suppose $\exists x : I \to \mathbb{R}^n$, diff, $x(I) \subset S$, $\exists t^* \in I : x(t^*) = x^*$, $\frac{dx}{dt}(t^*) = y$

Since $x(t) \in S$, $\forall t \in I$ we know $h(x(t)) = 0$. Take a derivative w.r.t. t

$$\frac{d}{dt} h(x(t)) = Dh(x(t)) \cdot \frac{dx}{dt}(t) = 0 \quad . \text{ For } t = t^*$$

$$Dh(x^*)\, y = 0 \quad \text{so } y \in T(x^*)$$

($\Rightarrow$) (skipped) requires the Implicit Function Theorem

- The *normal space* at x^* on the surface $S = \{x \mid h(x) = 0\}$

$N(x^*) = \{x \in \mathbb{R}^n \mid x = Dh(x^*)^T z,$ for some $z \in \mathbb{R}^m\}$ = Range or Image $Dh(x^*)^T \underset{\text{NOT}}{=} R(Dh(x^*)^T)$

Note that $0 \in N(x^*)$ and that $N(x^*)$ is spanned by the vectors $\{\nabla h_i(x^*)\}_{i=\overline{1,m}}$, that is,

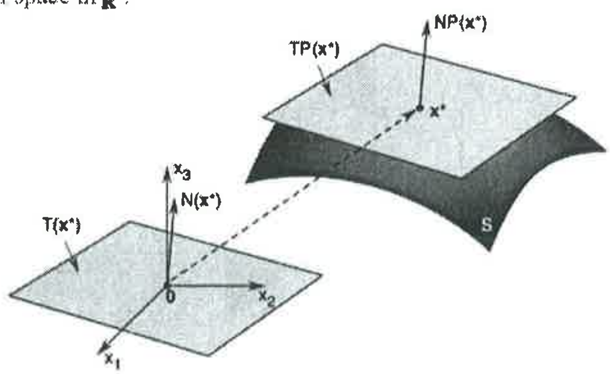
$x \in N(x^*) \Leftrightarrow \exists\, z_1, \dots, z_m \in \mathbb{R}$ s.t. $x = z_1 \nabla h_1(x^*) + \dots + z_m \nabla h_m(x^*)$

Assuming that x^* is regular, $\dim N(x^*) =$ rank $Dh(x^*) = m$

The *normal plane* at x^*

$$NP(x^*) = N(x^*) + x^* = \{x + x^* \mid x \in N(x^*)\}$$

Normal space in $\mathbb{R}^3$.



Lemma. $T(x^*) = N(x^*)^\perp$, $T(x^*)^\perp = N(x^*)^{\perp\perp} = N(x^*)$. In particular, every vector $v \in \mathbb{R}^n$ can be represented uniquely as $v = y + w$ with $y \in T(x^*)$, $w \in N(x^*)$ (or $\mathbb{R}^n = T(x^*) \oplus N(x^*)$).

Proof. "$\subset$" For every $y \in T(x^*)$, $x \in N(x^*)$ there is $z \in \mathbb{R}^m$ s.t. $x = Dh(x^*)^T z$

$x^T y = z^T \underbrace{Dh(x^*) y} = z^T \cdot 0 = 0 \Rightarrow T(x^*) \subset N(x^*)^\perp$. Conversely if $y \in N(x^*)^\perp$

by its def $\forall x \in N(x^*)$, $x^T y = 0$ $\Rightarrow z^T(Dh(x^*)^T y) = 0$, $\forall z \in \mathbb{R}^m \Rightarrow Dh(x^*)^T y = 0$

$\forall z \in \mathbb{R}^m$, $x = D(h(x^*)^T z$

that is $y \in T(x^*)$. Hence $T(x^*) = N(x^*)^\perp$

$T(x^*)^\perp = N(x^*)^{\perp\perp} = N(x^*)$ bc for every subspace $V \subset \mathbb{R}^n$, $V^{\perp\perp} = V$ $\supset$ trivial

$V \oplus V^\perp = \mathbb{R}^n$

Example.

#20.4 p 459

$S = \{x \in \mathbb{R}^3 \mid h_1(x) = x_1 = 0,\ h_2(x) = x_1 - x_2 = 0\}$

$Dh(x) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}$, $T(x) = \{y \mid \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_{y_1 = 0 \text{ and } y_1 - y_2 = 0 \Rightarrow y_2 = 0} = 0\} = \{[0, 0, \alpha]^T \mid \alpha \in \mathbb{R}\}$ - the x_3-axis

$N(x)$ = the $(x_1 x_2)$ plane!

(Necessary Cond) FONC

Theorem (Lagrange Multipliers) Let x^* be a local extrema of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $h(x) = 0$, where $h : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$. If x^* is a regular point then there exists $\lambda^* \in \mathbb{R}^m$ such that

$$Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T \text{ or } \nabla f(x^*) = Dh(x^*)^T(-\lambda^*)$$

Proof. Equivalently, $\nabla f(x^*) \in R(Dh(x^*)^T) =: N(x^*) = T(x^*)^\perp$

It suffices to prove that $\forall y \in T(x^*), \nabla f(x^*)^T y = 0$

Let $y \in T(x^*) \Rightarrow \exists\, x : I \to \mathbb{R}^n$ diff, $x(I) \subset S$, $\exists t^* \in I$ s.t. $x(t^*) = x^*$, $\frac{dx}{dt}(t^*) = y$

Take $\phi(t) = f(x(t))$. Then t^* is a local extrema for ϕ bc x^* is a local ext of f

So t^* is a critical number of ϕ, that is, $\frac{d\phi}{dt}(t^*) = Df(x^*) \cdot \underbrace{\frac{dx}{dt}(t^*)}_{y^*} = \underbrace{Df(x^*)}_{\nabla f(x^*)^T} y = 0$

$\nabla f(x^*)^T y = 0, \forall y \in T(x^*)$ so $\nabla f(x^*) \in T(x^*)^\perp$

QED

Lagrange's theorem states that if x^* is an extrema then the gradient of the objective function is a linear combination of the gradients of the constraints.

The vector λ^* is called the *Lagrange multiplier vector* and its components the *Lagrange multipliers.* A convenient way to apply Lagrange's theorem is provided by the Lagrangian function $L : \mathbb{R}^{m+n} \to \mathbb{R}$, defined by

$$L(x, \lambda) = f(x) + \lambda^T h(x).$$

Note that the unrestricted FONC for L: $0 = \nabla L(x, \lambda) = \begin{bmatrix} \nabla f(x) + Dh(x)^T \lambda \\ h(x) \end{bmatrix}$ is equivalent to the combination of the Lagrangian necessary condition and the constraint.

Example #20.5 p467 (P) Min $f(x) = x$ subject to $h(x) = 0$

$$h(x) = \begin{cases} x^2 & \text{if } x < 0 \\ 0 & \text{if } 0 \leq x \leq 1 \\ (x-1)^2 & \text{if } x > 1 \end{cases}$$

Is $h \in C^1(\mathbb{R})$? Yes

Feasible set $\{x \mid h(x) = 0\} = [0,1]$

$x^* = 0$ is a local min of (P)

However $f'(0) = 1$, $h'(0) = 0$ so $R(h'(0)) = \{0\}$ and clearly $f'(0) \notin R(h'(0))$

So $x^* = 0$ does not have Lagrange Multipliers! Why NOT?

Because $x^* = 0$ is NOT regular!

Example #20.8 p470 (P_0) Maximize $\dfrac{x^TQx}{x^TPx}$ on all $x \in \mathbb{R}^n$, $x \neq 0$,

where $Q = Q^T$, $Q \geq 0$ (positive semidefinite) $x^TQx \geq 0, \forall x \in \mathbb{R}^n$
$P = P^T$, $P > 0$ (positive definite) $x^TPx > 0, \forall x \in \mathbb{R}^n, x \neq 0$

x^* is a solution of (P_0) iff tx^* is a solution of (P_0) $\forall t \in \mathbb{R}\setminus\{0\}$

To avoid solution multiplicity we superimpose an extra cond $x^TPx = 1$

We get the equivalent problem

(P_1) Max x^TQx on all $x \in \mathbb{R}^n$ subject to $\underbrace{x^TPx = 1}$.

$$h(x) = 1 - x^TPx = 1 - \sum P_{ij} x_i x_j = 0$$

We prove first that any feasible point of (P_1) is REGULAR!

$\nabla h(x) = -2Px$ Assume (by contra) that x is non-regular (feasible)

Hence $Px = 0$ and $x^T\underbrace{Px}_{0} = 1 \Rightarrow 0 = 1$ (Contradiction)

$$L(x,\lambda) = x^TQx + \lambda(1 - x^TPx)$$

Assume x is a solution of $(P_1) \overset{FONC}{\Longrightarrow} \exists \lambda \in \mathbb{R}$, $DL(x,\lambda) = \begin{bmatrix} 2x^TQ - 2\lambda x^TP \\ 1 - x^TPx \end{bmatrix} = 0 \Rightarrow$

$Qx - \lambda Px = 0$ so $(Q - \lambda P)x = 0$ or $(\lambda P - Q)x = 0$

Since $P > 0$, Null$(P) = \{0\}$ so P^{-1} exists $\Rightarrow (\lambda I_n - P^{-1}Q)x = 0$

$P^{-1}Qx = \lambda x$, that is, λ is an eigenvalue of $P^{-1}Q$
x — " — eigenvector — " —

But $x^TPx = 1$ $\quad Qx = \lambda Px \,/\, \overset{\text{left}}{x^T} \Rightarrow x^TQx = \lambda x^TPx = \lambda$

If x^* is a solution of (P_1) then $\lambda^* = x^{*T}Qx^*$ is the largest eigenvalue of $P^{-1}Q$

Homework 6. Due: Name:

Show all your work and all details. Write accurately.

Problem 1. (a) Where is the assumption: "x^* is regular" essential in the proof of the results of section: Lagrange Multipliers

(b) In the example on page 49 (#20.8 in the book) explain in what way is (P_0) equivalent to (P_1). Prove your statements.

(c) State the SOSC Theorem on p. 51 (or Theorem 20.5 p. 474 in the book) for x^* a local maximizer.

Problem 2. #20.2(c) on page 482

Problem 3. #20.8 on page 483

Problem 4. #20.18 on page 485 (follow the class notes)

Problem 5. #20.21 on page 486 (match it to a quadratic programming problem)

Start every new problem solution on the top of the page.
Sign every sheet of paper you use.
Do not staple!

Homework 6 – Solutions

Total 80p

Problem 1. (a) Where is the assumption: "x^* is regular" essential in the proof of the results of section: Lagrange Multipliers

That assumption is essential in the Implicit Function Theorem which in turn is used for the proof of the theorem on p. 46 about the characterization of $y \in T(x^*)$ (iff there exists a differentiable curve in S passing through x^* with derivative y at x^*). 4

(b) In the example on page 49 (#20.8 in the book) explain in what way is (P_0) equivalent to (P_1). Prove your statements.

$$(P_0)\quad \text{Maximize } \frac{x^TQx}{x^TPx}, \text{ on all } x \in \mathbb{R}^n, x \neq 0, \quad \text{where } Q = Q^T \geq 0,\ P = P^T > 0. \qquad (P_1)\quad \text{Maximize } x^TQx \text{ subject to } x^TPx = 1.$$

They are equivalent in the following sense:

1. If x^* is a solution of (P_1) then, for every $t \neq 0$, tx^* is a solution of (P_0). 8
2. If x^* is a solution of (P_0) then $y = \dfrac{1}{\sqrt{x^{*T}Px^*}}x^*$ is a solution of (P_1). 8

1. Let x^* be a solution of (P_1), that is, $x^{*T}Px^* = 1$ and, for every u such that $u^TPu = 1$, $u^TQu \leq x^{*T}Qx^*$ (*).
   For every $t \neq 0$ denote by $x_t = tx^*$. Note that $\dfrac{x_t^TQx_t}{x_t^TPx_t} = \dfrac{x^{*T}Qx^*}{x^{*T}Px^*} = x^{*T}Qx^*$.
   For every $x \in \mathbb{R}^n$, $x \neq 0$, take $u = \dfrac{1}{\sqrt{x^TPx}}x$. Then $u^TPu = 1$ and we can use (*) to get $\dfrac{x^TQx}{x^TPx} = u^TQu \leq x^{*T}Qx^* = \dfrac{x_t^TQx_t}{x_t^TPx_t}$, that is, x_t is a solution of (P_0).

2. Let x^* be a solution of (P_0), that is, for every $v \in \mathbb{R}^n$, $v \neq 0$, $\dfrac{v^TQv}{v^TPv} \leq \dfrac{x^{*T}Qx^*}{x^{*T}Px^*} = y^TQy$ (**).
   For every x such that $x^TPx = 1$, we use (**) to get $x^TQx = \dfrac{x^TQx}{x^TPx} \leq y^TQy$ and since $y^TPy = 1$ we get that y is a solution of (P_1).

(c) State the SOSC Theorem on p. 51 (or Theorem 20.5 p. 474 in the book) for x^* a local maximizer.

Theorem (SOSC-max). Suppose that $f, h \in C^2$ and there exist $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$ such that: 4

1. $Df(x^*) + \lambda^{*T}Dh(x^*) = 0^T$,
2. For all $y \in T(x^*)$, $y \neq 0$, we have $y^TD_x^2L(x^*, \lambda^*)y < 0$.

Then x^* is a strict local maximizer of f subject to $h(x) = 0$.

Problem 2. #20.2(c) on page 482 Find local extremizers for the following optimization problem: Maximize $x_1 x_2$, subject to $x_1^2 + 4x_2^2 = 1$.

$L(x_1, x_2, \lambda) = x_1 x_2 + \lambda(x_1^2 + 4x_2^2 - 1)$. The Lagrange-FONC are

$$\frac{\partial L}{\partial x_1} = x_2 + 2\lambda x_1 = 0$$
$$\frac{\partial L}{\partial x_2} = x_1 + 8\lambda x_2 = 0$$
$$x_1^2 + 4x_2^2 = 1.$$

If $x_1 = 0$ or $x_2 = 0$ then $x_1 = x_2 = 0$ in contradiction with the 3rd equation.

Hence $x_1 \neq 0$, $x_2 \neq 0$ and $\frac{x_2}{x_1} = -2\lambda$, $\frac{x_1}{x_2} = -8\lambda \Rightarrow 16\lambda^2 = 1$ so $\lambda = \pm\frac{1}{4}$.

For $\lambda = 1/4$ we get $x_1 = -2x_2$ and from the 3rd $8x_2^2 = 1 \Rightarrow x_2 = \pm\frac{1}{2\sqrt{2}}$ and the points $(-\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}})$, $(\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}})$.

For $\lambda = -1/4$ we get $x_1 = 2x_2$ and from the 3rd $8x_2^2 = 1 \Rightarrow x_2 = \pm\frac{1}{2\sqrt{2}}$ and the points $(\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}})$, $(-\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}})$.

$h(x_1, x_2) = x_1^2 + 4x_2^2 - 1$, $\nabla h(x_1, x_2) = (2x_1, 8x_2)^T$ so all the above points are regular ($\nabla h \neq 0$).

8 $D_x^2 L(x_1, x_2, \lambda) = \begin{bmatrix} 2\lambda & 1 \\ 1 & 8\lambda \end{bmatrix}$

8 For $(-\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}, \frac{1}{4})$, $D_x^2 L = \begin{bmatrix} 1/2 & 1 \\ 1 & 2 \end{bmatrix}$, $Dh(x)y = [-\sqrt{2}, 2\sqrt{2}]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \Rightarrow y_1 = 2y_2$ so $T(-\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}) = \{\begin{bmatrix} 2a \\ a \end{bmatrix} \mid a \in \mathbb{R}\}$, $[2a, a]\begin{bmatrix} 1/2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2a \\ a \end{bmatrix} = 8a^2 > 0$, for $a \neq 0$ so $(-\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}})$ is a strict local minimizer.

For $(\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}}, \frac{1}{4})$, $D_x^2 L = \begin{bmatrix} 1/2 & 1 \\ 1 & 2 \end{bmatrix}$, $Dh(x)y = [\sqrt{2}, -2\sqrt{2}]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \Rightarrow y_1 = 2y_2$ so $T(\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}}) = \{\begin{bmatrix} 2a \\ a \end{bmatrix} \mid a \in \mathbb{R}\}$, $[2a, a]\begin{bmatrix} 1/2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2a \\ a \end{bmatrix} = 8a^2 > 0$, for $a \neq 0$ so $(\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}})$ is a strict local minimizer.

For $(\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}, -\frac{1}{4})$, $D_x^2 L = \begin{bmatrix} -1/2 & 1 \\ 1 & -2 \end{bmatrix}$, $Dh(x)y = [\sqrt{2}, 2\sqrt{2}]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \Rightarrow y_1 = -2y_2$ so $T(\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}) = \{\begin{bmatrix} -2a \\ a \end{bmatrix} \mid a \in \mathbb{R}\}$, $[-2a, a]\begin{bmatrix} -1/2 & 1 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} -2a \\ a \end{bmatrix} = -8a^2 < 0$, for $a \neq 0$ so $(\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}})$ is a strict local maximizer.

For $(-\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}}, -\frac{1}{4})$, $D_x^2 L = \begin{bmatrix} -1/2 & 1 \\ 1 & -2 \end{bmatrix}$, $Dh(x)y = [-\sqrt{2}, -2\sqrt{2}]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \Rightarrow y_1 = -2y_2$ so $T(-\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}}) = \{\begin{bmatrix} -2a \\ a \end{bmatrix} \mid a \in \mathbb{R}\}$, $[-2a, a]\begin{bmatrix} -1/2 & 1 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} -2a \\ a \end{bmatrix} = -8a^2 < 0$, for $a \neq 0$ so $(-\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}})$ is a strict local maximizer.

Problem 3. #20.8 on page 483 Consider the problem: Minimize $2x_1 + 3x_2 - 4$, $x_1, x_2 \in \mathbb{R}$, subject to $x_1 x_2 = 6$.

a. Use Lagrange's theorem to find all possible local minimizers and maximizers. 8

b. Use the second-order sufficient conditions to specify which points are strict local minimizers and which are strict local maximizers. 8

c. Are the points in part b global minimizers or maximizers? Explain. 4

a. $f(x_1, x_2) = 2x_1 + 3x_2 - 4$, $h(x_1, x_2) = x_1x_2 - 6$, $Df(x_1, x_2) = [2, 3]$, $Dh(x_1, x_2) = [x_2, x_1]$.
Note that $(0, 0)$ is not a feasible point. Therefore, any feasible point is regular.
$L(x_1, x_2, \lambda) = 2x_1 + 3x_2 - 4 + \lambda(x_1x_2 - 6)$. The Lagrange-FONC are

$$\frac{\partial L}{\partial x_1} = 2 + \lambda x_2 = 0$$
$$\frac{\partial L}{\partial x_2} = 3 + \lambda x_1 = 0$$
$$x_1 x_2 = 6.$$

Clearly $\lambda \neq 0$, $x_1 = -3/\lambda$, $x_2 = -2/\lambda$ so $x_1x_2 = 6/\lambda^2 = 6 \Rightarrow \lambda = \pm 1$, and the possible local minimizers and maximizers are $(-3, -2)$, $(3, 2)$.

b. $D_x^2 L(x_1, x_2, \lambda) = \begin{bmatrix} 0 & \lambda \\ \lambda & 0 \end{bmatrix}$

$(-3, -2, 1) \Rightarrow D_x^2 L(x_1, x_2, \lambda) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $Dh(x)y = [-2, -3]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \Rightarrow 2y_1 + 3y_2 = 0$ so $T(-3, -2) = \{[-3a, 2a]^T \mid a \in \mathbb{R}\}$ $[-3a, 2a]\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -3a \\ 2a \end{bmatrix} = -12a^2 < 0$, for $a \neq 0$ so by SOSC $(-3, -2)$ is a strict local maximizer.

$(3, 2, -1) \Rightarrow D_x^2 L(x_1, x_2, \lambda) = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$, $Dh(x)y = [2, 3]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \Rightarrow 2y_1 + 3y_2 = 0$ so $T(3, 2) = \{[-3a, 2a]^T \mid a \in \mathbb{R}\}$ $[-3a, 2a]\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} -3a \\ 2a \end{bmatrix} = 12a^2 > 0$, for $a \neq 0$ so by SOSC $(3, 2)$ is a strict local minimizer.

c. $f(-3, -2) = -16 < 8 = f(3, 2)$ so both points are not global extrema.

Problem 4. #20.18 on page 485 (follow the class notes) Consider the problem of minimizing a general quadratic function subject to a linear constraint:

8

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx - c^Tx + d \\ \text{subject to} & Ax = b, \end{array}$$

where $Q = Q^T > 0$, $A \in \mathbb{R}^{m \times n}$ with $m < n$, rank$A = m$ and d is a constant. Derive a closed form solution to the problem

$L(x, \lambda) = \frac{1}{2}x^TQx - c^Tx + d + \lambda(b - Ax)$. Suppose that x^* is a solution of the problem. Since the rank of $Dh(x^*) = -A$ is m we know that x^* is regular . According to Lagrange-FONC, there exits λ^* such that $D_xL(x, \lambda) = x^{*T}Q - c^T - \lambda^{*T}A = 0$ so $Qx^* = c + A^T\lambda^* \Rightarrow x^* = Q^{-1}A^T\lambda^* + Q^{-1}c$ (*)

But $Ax^* = b$ gives $AQ^{-1}A^T\lambda + AQ^{-1}c = b$ so $\lambda^* = (AQ^{-1}A^T)^{-1}b - (AQ^{-1}A^T)^{-1}AQ^{-1}c$ and from (*) we get

$$x^* = Q^{-1}A^T(AQ^{-1}A^T)^{-1}b - Q^{-1}A^T(AQ^{-1}A^T)^{-1}AQ^{-1}c + Q^{-1}c.$$

Problem 5. #20.21 on page 486 (match it to a quadratic programming problem) Consider the discrete-time linear system $x_k = 2x_{k-1} + u_k$, $k \geq 1$, with $x_0 = 1$. Find the values of the control inputs u_1 and u_2 to minimize

$$x_2^2 + \frac{1}{2}u_1^2 + \frac{1}{3}u_2^2.$$

Letting $z = [x_2, u_1, u_2]^T$ then $x_2^2 + \frac{1}{2}u_1^2 + \frac{1}{3}u_2^2 = z^TQz$, where $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$. > 0

The linear constraint on z is obtained by writing $x_2 = 2x_1 + u_2 = 2(2 + u_1) + u_2 = 2u_1 + u_2 + 4$ and seeing it as $Az = b$ where $A = [1, -2, -1]$, $b = 4$.

6

6

This is a quadratic programming problem with solution

$$z^* = Q^{-1}A^T(AQ^{-1}A^T)^{-1}b = [1/3, -4/3, -1]^T \Rightarrow u_1^* = -4/3, \; u_2^* = -1.$$

5.1.1 Second-Order Conditions for (NE)

We assume that $f \in C^2(\mathbb{R}^n)$ and $h \in C^2(\mathbb{R}^n; \mathbb{R}^m)$, $h = (h_1, \ldots, h_m)$.

We denote $D^2 h_k(x) = H_k(x) = (\frac{\partial^2 h_k}{\partial x_i \partial x_j})_{i,j=\overline{1,n}}$ the Hessian matrix of h_k, $k = \overline{1, m}$ and by $D_x^2 L$ the Hessian matrix of $L(x, \lambda)$ with respect to x. We use the notation

$$\lambda \cdot H(x) = \lambda_1 H_1(x) + \ldots + \lambda_m H_m(x).$$

Theorem. (SONC) Let x^* be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $h(x) = 0$, where $h : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$, and $f, h \in C^2$. Suppose that x^* is regular. Then, there exists λ^* such that:

1. $Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$,
2. For all $y \in T(x^*)$, we have $y^T D_x^2 L(x^*, \lambda^*) y \geq 0$.

Proof. Condition 1 is FONC (see Lagrange's Thm, p.48) It remains to prove 2.

Let $y \in T(x^*) \Rightarrow \exists x : I \to \mathbb{R}^n$, $x \in C^2$, $x(I) \subset S = \{h = 0\}$, $\exists t^* \in I$, $x(t^*) = x^*$, $\frac{dx}{dt}(t^*) = y$

Again t^* is a local min of $\phi(t) = f(x(t))$ so ϕ is concave up (convex) around t^* so $\frac{d^2\phi}{dt^2}(t^*) \geq 0$.

$\frac{d\phi}{dt}(t) = Df(x(t)) \cdot \frac{dx}{dt}(t)$ (chain rule)

$$\frac{d^2\phi}{dt^2} = \frac{d}{dt}\left(Df(x(t)) \cdot \frac{dx}{dt}(t)\right) = \frac{dx}{dt}(t)^T D^2 f(x(t)) \cdot \frac{dx}{dt}(t) + Df(x(t)) \cdot \frac{d^2x}{dt^2}(t)$$

For $t = t^*$ we get $y^T D^2 f(x^*) y + Df(x^*) \frac{d^2x}{dt^2}(t^*) \geq 0$ $(*)$

We have $x(I) \subset S$, i.e., $h(x(t)) = 0, \forall t \in I \Rightarrow \lambda^* \cdot h(x(t)) = 0 \Rightarrow \frac{d^2}{dt^2}(\lambda^* \cdot h(x(t))) = 0$

$$\frac{d^2}{dt^2}(\lambda^* \cdot h(x(t))) = \frac{d}{dt}\left[\lambda^* \cdot \frac{d}{dt} h(x(t))\right] = \frac{d}{dt} \sum_{k=1}^m \lambda_k^* \underbrace{\frac{d}{dt} h_k(x(t))}_{Dh_k(x(t)) \cdot \frac{dx}{dt}} = \sum_{k=1}^m \lambda_k^* \frac{d}{dt}\left[Dh_k(x(t)) \cdot \frac{dx}{dt}(t)\right]$$

$$= \sum_{k=1}^m \lambda_k^* \left[\frac{dx}{dt}^T D^2 h_k(x(t)) \cdot \frac{dx}{dt} + Dh_k(x(t)) \cdot \frac{d^2x}{dt^2}(t)\right] = \frac{dx}{dt}^T \left[\lambda^* \cdot D^2 h(x(t))\right] \frac{dx}{dt} + \lambda^* \cdot Dh(x(t)) \frac{d^2x}{dt^2} = 0$$

For $t = t^*$ we get $y^T [\lambda^* \cdot D^2 h(x^*)] y + \lambda^{*T} Dh(x^*) \cdot \frac{d^2x}{dt^2}(t^*) = 0$ $(**)$

$$(*) + (**) \quad y^T \underbrace{[D^2 f(x^*) + \lambda^* \cdot D^2 h(x^*)]}_{D_x^2 L(x^*, \lambda^*)} y + \underbrace{[Df(x^*) + \lambda^{*T} Dh(x^*)]}_{= 0 \text{ bc of } 1.} \frac{d^2x}{dt^2}(t^*) \geq 0$$

$L(x^*, \lambda^*) = f(x^*) + \lambda^* \cdot h(x)$ So 2.

Theorem. (SOSC) Suppose that $f, h \in C^2$ and there exist $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$ such that:

1. $Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$,
2. For all $y \in T(x^*)$, $y \neq 0$, we have $y^T D_x^2 L(x^*, \lambda^*) y > 0$. (skipped proof)

Then x^* is a strict local minimizer of f subject to $h(x) = 0$, that is, $\forall x$ s.t. $h(x)=0, x \neq x^*$, $f(x^*) < f(x)$

Interpretation: If x^* is a local min of (NE) then $D_x^2 L(x^*,\lambda^*)$ is positive semidefinite on $T(x^*)$. Conversely, if $D_x^2 L(x^*,\lambda^*)$ is positive definite on $T(x^*)$ then x^* is a STRICT local min of (NE)

Example. 20.9 on p.475 Max $\dfrac{x^TQx}{x^TPx}$ on all $x \in \mathbb{R}^2, x \neq 0 \Rightarrow$ Max $\overbrace{x^TQx}^{f(x)}$ subj to $x^TPx = 1$

where $Q = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$, $P = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

$h(x) = 1 - x^TPx$

$Dh(x) = -2Px$

$L(x,\lambda) = x^TQx + \lambda(1 - x^TPx)$

Lagrange cond (FONC, p.49) $(\lambda I - P^{-1}Q)x = 0$ $\quad P^{-1}Q = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ eigenv: $\lambda_1 = 2$, $\lambda_2 = 1$

So $\lambda^* = \lambda_1 = 2$ $\quad 2I - \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{matrix} x_1 \\ x_2 \end{matrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$ so $x_2 = 0$

$E_{\lambda^*} = \left\{ \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \middle| \alpha \in \mathbb{R} \right\}$ $\quad [\alpha \; 0]\underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}}_{P}\begin{bmatrix} \alpha \\ 0 \end{bmatrix} = 2\alpha^2 = 1, \alpha = \pm\frac{1}{\sqrt{2}}$

Denote by $x^* = [\frac{1}{\sqrt{2}}, 0]^T$. Eigenvectors are $\pm x^*$ They are both local maxs!

$D_x^2 L(x^*,\lambda^*) = 2Q - 2\lambda^* P = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 8 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}$

$T(x^*) = \{ y \in \mathbb{R}^2 \mid Dh(x^*)y = \underbrace{-2x^{*T}P}_{[-2\sqrt{2}, 0]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}} y = 0 \} = \{ y \mid y_1 = 0 \} = \{ y = \begin{bmatrix} 0 \\ a \end{bmatrix} \mid a \in \mathbb{R} \}$

2. $\forall y \in T(x^*), y \neq 0$ (that is $a \neq 0$) $\quad y^T D_x^2 L(x^*,\lambda^*) y = [0, a]\begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} 0 \\ a \end{bmatrix} = -2a^2 < 0$

So x^* is a local MAX bc $D_x^2 L$ is negative definite (on $T(x^*)$)

Similarly for $-x^*$.

5.1.2 Quadratic Programming I

#22.12

$$\text{(Q) Minimize } \frac{1}{2}x^T Q x \text{ subject to } Ax = b,$$

where $Q = Q^T > 0$, that is, Q is positive definite: for every $x \in \mathbb{R}^n$, $x \neq 0$, $x^T Q x > 0$, $h(x) = b - Ax$
$A \in \mathbb{R}^{m\times n}$, $m < n$, rank $A = m$.

$L(x,\lambda) = \frac{1}{2}x^T Q x + \lambda^T(b - Ax)$

Lagrange Cond (FONC) for sol x^* and λ^* : $D_x L(x^*,\lambda^*) = x^{*T}Q - \lambda^{*T}A = 0 \; /^T$

$Qx^* = A^T\lambda^* \Rightarrow x^* = Q^{-1}A^T\lambda^*$ (1) $\quad Ax^* = b \overset{(1)}{\Rightarrow} (AQ^{-1}A^T)\lambda^* = b$ (2)

But $AQ^{-1}A^T$ is positive definite so invertible

$x^T(AQ^{-1}A^T)x = (A^Tx)^T Q^{-1} \underbrace{(A^Tx)}_{z} = z^T Q^{-1} z > 0, \; \forall x \neq 0$

$Q^{-1} > 0$ $\quad$ bc $A^Tx = 0 \Rightarrow x = 0$
rank $A = m$
$(x \neq 0 \Rightarrow z \neq 0)$

$(2) \Rightarrow \lambda^* = (AQ^{-1}A^T)^{-1}b \Rightarrow x^* = Q^{-1}A^T(AQ^{-1}A^T)^{-1}b$
the only candidate for a minimizer

$D_x^2 L(x^*,\lambda^*) = Q > 0$ everywhere (not only on $T(x^*)$!) $\overset{\text{SOSC}}{\Longrightarrow}$ x^* is a local min

Because f and h are convex $\underset{\text{later}}{\Longrightarrow}$ x^* is a global min

<u>Particular Case</u> $\quad Q = I_n$, $f(x) = \frac{1}{2}x^TQx = \frac{1}{2}\|x\|^2 = \frac{1}{2}\sum_{i=1}^n x_i^2$ s.t. $Ax = b$

The global min is $x^* = A^T(AA^T)^{-1}b$

Example. #20.10 p478 $x_k = a x_{k-1} + b u_k \ , k \geq 1$

$\underbrace{\text{Min } \frac{1}{2} \sum_{i=1}^{n} q x_i^2 + r u_i^2}_{\text{weighted sum approach}}$ subject to ↑ $, k = \overline{1, N}$ x_0 - given

$$z = \begin{bmatrix} x \\ u \end{bmatrix}_{2N\times 1}, \quad Q = \begin{bmatrix} qI_N & 0 \\ 0 & rI_N \end{bmatrix}, \quad A = \left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & -b & 0 & \cdots & 0 \\ -a & 1 & \cdots & 0 & 0 & -b & \cdots & 0 \\ & -a & & & & & & \\ 0 & & -a & 1 & 0 & 0 & \cdots & -b \end{array}\right]^{2N\times 2N}$$

$$b = \begin{bmatrix} a x_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow \text{Min } \frac{1}{2} z^T Q z \text{ subj to } Az = b \text{ with solution}$$

$$z^* = Q^{-1} A^T (A Q^{-1} A^T)^{-1} b$$

Example. #20.11

$x_0 = 10{,}000$ x_k - account balance at the end of month k

$a = 1.02$ u_k - payment in month k

$k = \overline{1, 10}$ $N = 10$

(SE) $x_k = 1.02 x_{k-1} - u_k$ so $b = -1$

$\text{Min } \frac{1}{2} \sum_{i=1}^{10} q x_i^2 + r u_i^2$ subj to (SE)

$\frac{q}{r}$ large → reduce debt

$\frac{r}{q}$ large → reluctance to pay

$(q = 1)$

$q = 1, r = 10$

Example. #20.2(b) p.482 Max $4x_1+x_2^2$ s.t. $x_1^2+x_2^2=9$

$h(x_1,x_2)=x_1^2+x_2^2-9=0$

$L(x_1,x_2,\lambda)=4x_1+x_2^2+\lambda(x_1^2+x_2^2-9)$, $\nabla L=0 \Rightarrow \begin{cases} \frac{\partial L}{\partial x_1} = 4+2\lambda x_1=0 \rightarrow x_1=-\frac{2}{\lambda} \\ 2x_2+2\lambda x_2=0 \xrightarrow{2^{nd}} \\ x_1^2+x_2^2=9 \end{cases}$

$2x_2(1+\lambda)=0$ → $x_2=0$ or $\lambda=-1$

$x_2=0 \xrightarrow{3^{rd}} x_1=\pm 3$ $\quad(3,0,-\frac{2}{3}),(-3,0,\frac{2}{3})$

$\lambda=-1 \xrightarrow{1^{st}} x_1=2 \xrightarrow{3^{rd}} x_2=\pm\sqrt{5}$, $(2,\pm\sqrt{5},-1)$

$\nabla h(x_1,x_2)=2[x_1,x_2]^T\neq 0$ for all c.p. so they are all regular

$D_x^2L(x_1,x_2,\lambda)=\begin{bmatrix}2\lambda & 0\\ 0 & 2+2\lambda\end{bmatrix}$ Check SOSC

For $(3,0,-\frac{2}{3})$, $D_x^2L(3,0,-\frac{2}{3})=\begin{bmatrix}-4/3 & 0\\ 0 & 2/3\end{bmatrix}$

$Dh(x^*)y=2[3,0]\begin{bmatrix}y_1\\y_2\end{bmatrix}=0 \Rightarrow y_1=0 \Rightarrow T(3,0)=\left\{\begin{bmatrix}0\\a\end{bmatrix} \mid a\in\mathbb{R}\right\}$

$y^T\cdot D_x^2L\,y=[0,a]\begin{bmatrix}-4/3 & 0\\ 0 & 2/3\end{bmatrix}\begin{bmatrix}0\\a\end{bmatrix}=\frac{2}{3}a^2>0, \forall a\neq 0 \Rightarrow (3,0)$ is a strict local min

For $(-3,0,\frac{2}{3}) \Rightarrow D_x^2L(-3,0,\frac{2}{3})=\begin{bmatrix}4/3 & 0\\ 0 & 10/3\end{bmatrix}>0 \Rightarrow (-3,0)$ strict local min

$(2,\pm\sqrt{5})$ are strict local max's

5.2 Karush-Kuhn Tucker Condition

(N) Minimize (Maximize) $f(x)$ subject to $h(x) = 0$ and $g(x) \le 0$,

$h: \mathbb{R}^n \to \mathbb{R}^m$, $g: \mathbb{R}^n \to \mathbb{R}^p$, and $m+p \le n$!

More notions associated to (N)

- An inequality constraint $g_j(x) \le 0$ is said to be *active* at x^* if $g_j(x^*)=0$. It is *inactive* at x^* if $g_j(x^*)<0$.
We denote *the index set of active inequality constraints* by

$$J(x^*) := \{ j \in \overline{1,p} \mid g_j(x^*) = 0 \}$$

- $x^* \in \mathbb{R}^n$ is *regular* if the vectors
$\{ \nabla h_i(x^*), \nabla g_j(x^*) \mid i=\overline{1,m}, j \in J(x^*) \}$ are linearly independent

5.2.1 First-Order Necessary Conditions for (N)

Theorem (KKT-FONC). Let $f, h, g \in C^1$. Let x^* be a regular point and a local minimizer for the problem (N-min): Minimize f subject to $h(x) = 0$, $g(x) \le 0$. Then, there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that:

1. $\mu^* \ge 0$. i.e. $\mu_j^* \ge 0, \forall j = \overline{1,p}$
2. $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$ or $\nabla f(x^*) + \nabla h(x^*) \lambda^* + \nabla g(x^*) \mu^* = 0$ (transpose)
3. $\mu^{*T} g(x^*) = 0$. or $0 = \sum_{j=1}^{p} \mu_j^* g_j(x^*) = \sum_{j \notin J(x^*)} \mu_j^* g_j(x^*) \overset{1.}{\Longleftrightarrow} \forall j \notin J(x^*), \mu_j^* = 0$

Proof. Let $S = \{ x \mid h(x) = 0, g(x) \le 0 \}$ be the feasible set
$S' := \{ x \mid h(x) = 0, \forall j \in J(x^*), g_j(x) = 0 \}$, where x^* is the regular local min of f over S; $x^* \in S'$ due to the def of $J(x^*)$. Goal: x^* is a local min of f over S'!
Since x^* is a local min of f over S, there is B^* (an open ball containing x^*) such that

$$\forall x \in S \cap B^*, f(x) \ge f(x^*) \quad (1)$$

From the def of $J(x^*)$, $\forall j \notin J(x^*), g_j(x^*) < 0 \xRightarrow{g_j \text{ cont}} \exists B$ (open ball containing x^*) such that

$$\forall x \in B, \forall j \notin J(x^*), g_j(x) < 0 \quad (2)$$

For every $x \in S' \cap B^* \cap B$: bc $x \in B \overset{(2)}{\Rightarrow} \forall j \notin J(x^*), g_j(x) < 0$; bc $x \in S' \Rightarrow \forall j \in J(x^*), g_j(x) = 0$ $\Rightarrow \forall j \in \overline{1,p}, g_j(x) \le 0$; $g(x) \le 0$ also $h(x) = 0$ $\Rightarrow x \in S$
Since $x \in S \cap B^* \overset{(1)}{\Rightarrow} f(x) \ge f(x^*)$ So my goal is met
By Lagrange FONC p.48 (x^* is regular) $\exists \lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^p$ s.t.
$Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$, where $\forall j \notin J(x^*)$ we set $\mu_j^* = 0 \overset{(3)}{\Longrightarrow}$ 2. and 3. hold!
It remains to prove 1. $\Leftrightarrow \forall j \in J(x^*), \mu_j^* \ge 0$ (4) - proof by contradiction - skipped
(we assume $\exists j_0 \in J(x^*), \mu_{j_0}^* < 0$ and prove that x^* is NOT a local min of f on S)

Interpretation of KKT conditions: In problems we from the Lagrangian

$L(x,\lambda,\mu) := f(x) + \lambda \cdot h(x) + \mu \cdot g(x)$ Then 2. says $D_x L(x,\lambda,\mu) = 0$ (2)

We also have $\mu \geq 0$ (1) and $\mu \cdot g(x^*) = 0$ (3) but also $\begin{cases} h(x^*) = 0 & (4) \\ g(x^*) \leq 0 & (5) \end{cases}$

3 equalities and 2 ineq

Strategy! $\mu = 0$ together with $D_x L = 0$ and (4),(5) or

$\mu \neq 0 \Rightarrow g(x^*) = 0$

Minimize $x_1^2 + 4x_2^2$ subject to $x_1^2 + 2x_2^2 \geq 4 \Leftrightarrow g \leq 0$

Example #21.1 a. p. 501

so $g(x_1,x_2) = 4 - x_1^2 - 2x_2^2$

(a) $L(x_1,x_2,\mu) = x_1^2 + 4x_2^2 + \mu(4 - x_1^2 - 2x_2^2)$

$D_x L(x_1,x_2,\mu) = [2x_1 - 2\mu x_1, 8x_2 - 4\mu x_2] = 0$ $\Rightarrow$ $2x_1(1-\mu) = 0$, $4x_2(2-\mu) = 0$ so

$(x_1 = 0 \text{ or } \mu = 1)$ and $(x_2 = 0 \text{ or } \mu = 2)$ Also we $\mu \geq 0$ (1), $\mu(4 - x_1^2 - 2x_2^2) = 0$ (3)

$x_1^2 + 2x_2^2 \geq 4$, (5)

case (A) $\mu \neq 1, 2 \Rightarrow x_1 = x_2 = 0$ impossible due (5)

Case (B) $\mu = 1 \Rightarrow x_2 = 0 \underset{(3)}{\Rightarrow} x_1^2 = 4$ so $x_1 = \pm 2$ so $x = [\pm 2, 0]^T$

Case (C) $\mu = 2 \Rightarrow x_1 = 0 \underset{(3)}{\Rightarrow} x_2^2 = 2$ so $x_2 = \pm\sqrt{2}$ so $x = [0, \pm\sqrt{2}]^T$

$\nabla g(x_1,x_2) = \begin{bmatrix} -2x_1 \\ -4x_2 \end{bmatrix} \neq 0$ for all x's in cases (B),(C)

They are all regular

5.2.2 Second-Order Conditions for (N)

In this section, for a feasible point x^*, we denote by $T(x^*)$ the tangent space at x^* to the surface defined by the equality constraints $h(x) = 0$ and by the equations corresponding to active inequality constraints at x^*, namely, $g_j(x) = 0$, $j \in J(x^*)$:

$$T(x^*) = \{y \in \mathbb{R}^n \mid Dh(x^*)y = 0,\ \forall j \in J(x^*),\ Dg_j(x^*)y = 0\}.$$

Theorem (KKT-SONC). Let $f, h, g \in C^2$. Let x^* be a regular point and a local minimizer for the problem (N-min): Minimize f subject to $h(x) = 0$, $g(x) \leq 0$. Then, there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that:

1. $\mu^* \geq 0$, $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$, $\mu^{*T} g(x^*) = 0$. KKT-FONC
2. For all $y \in T(x^*)$, $y^T \underbrace{D_x^2[f(x^*) + h(x^*)^T\lambda^* + g(x^*)^T\mu^*]}_{D_x^2 L(x^*, \lambda^*, \mu^*)} y \geq 0$.

Proof.

Part 1. is exactly KKT-FONC

For Part 2. Recall that we proved that since x^* is a local min of f over $S = \{x \mid h(x) = 0, g(x) \leq 0\}$, x^* is also a local min of f over $S' = \{x \mid h(x) = 0, \forall j \in J(x^*), g_j(x) = 0\}$ (proved on p. 55)

We can use Lagrange SONC p. 50 to get 2.

QED

Let x^* be feasible for (N-min) and let $\mu^* \in \mathbb{R}^p$. Denote by $\tilde{J}(x^*, \mu^*) := \{j = \overline{1,p} \mid g_j(x^*) = 0,\ \mu_j^* > 0\}$ the index set of active inequality constraints (at x^*) for which $\mu_j^* > 0$. Clearly, $\tilde{J}(x^*, \mu^*) \subset J(x^*)$

Correspondingly we have the tangent space

$$\tilde{T}(x^*, \mu^*) = \{y \in \mathbb{R}^n \mid Dh(x^*)y = 0,\ \forall j \in \tilde{J}(x^*, \mu^*),\ Dg_j(x^*)y = 0\}.$$

to the surface defined by the equality constraints and the equations corresponding to active inequality constraints with $\mu_j^* > 0$. Note that the latter surface is possibly subject to fewer constraints, so $T(x^*)$ is a SUBSPACE of $\tilde{T}(x^*, \mu^*)$.

Theorem (KKT-SOSC). Let $f, h, g \in C^2$. Suppose that there exist a regular feasible point x^* of (N-min) and $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}^p$ such that:

1. $\mu^* \geq 0$, $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$, $\mu^{*T} g(x^*) = 0$.
2. For all $y \in \tilde{T}(x^*, \mu^*)$, $y \neq 0$, $y^T \underbrace{D_x^2[f(x^*) + h(x^*)^T\lambda^* + g(x^*)^T\mu^*]}_{D_x^2 L} y > 0$.

Then $x^*$is a strict local minimizer of (N-min). [Skipped Proof]

Example #21.1 b. p. 501 $D_x^2 L(x_1,x_2,\mu) = \begin{bmatrix} 2-2\mu & 0 \\ 0 & 8-4\mu \end{bmatrix}$

For $\mu=1$, $x=\begin{bmatrix}\pm 2\\ 0\end{bmatrix}$, $D_x^2 L(x,1) = \begin{bmatrix} 0 & 0\\ 0 & 4\end{bmatrix}$, $\tilde{J}(x,1)=\{1\}$

$\tilde{T}(x,1) = \{y \mid Dg(x^*)y=0\} = \{y \mid [\mp 4, 0]\begin{bmatrix}y_1\\ y_2\end{bmatrix}=0\} = \left\{\begin{bmatrix}0\\ a\end{bmatrix} \,\middle|\, a\in\mathbb{R}\right\}$ $y_1=0$

$g(x_1,x_2)=4-x_1^2-2x_2^2$, $Dg(x_1,x_2)=[-2x_1,-4x_2]$

$[a,a]\begin{bmatrix}0 & 0\\ 0 & 4\end{bmatrix}\begin{bmatrix}0\\ a\end{bmatrix} = 4a^2>0$, $\forall a\neq 0$ $(y\neq 0)$ so $\begin{bmatrix}\pm 2\\ 0\end{bmatrix}$ are strict local minimizers

For $\mu=2$, $x=\begin{bmatrix}0\\ \pm\sqrt{2}\end{bmatrix}$, $D_x^2 L(x,2) = \begin{bmatrix}-2 & 0\\ 0 & 0\end{bmatrix}$, $\tilde{J}(x,2)=\{1\}$

$\tilde{T}(x,2) = \{y \mid Dg(x^*)y=0\} = \left\{\begin{bmatrix}a\\ 0\end{bmatrix} \,\middle|\, a\in\mathbb{R}\right\}$

$[0, \mp 4\sqrt{2}]\begin{bmatrix}y_1\\ y_2\end{bmatrix}$

$y_2=0$

$[a,0]\begin{bmatrix}-2 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}a\\ 0\end{bmatrix} = -2a^2<0$ so $\begin{bmatrix}0\\ \pm\sqrt{2}\end{bmatrix}$ are NOT minimizers bc they don't satisfy KKT-SONC

Example 21.5 p.499 Min $x_1 x_2$ subj to $\begin{cases} x_1+x_2\geq 2 \to g_1(x_1,x_2)=2-x_1-x_2\leq 0 \\ x_2\geq x_1 \to g_2(x_1,x_2)=x_1-x_2\leq 0\end{cases}$

(a) KKT cond. $L(x_1,x_2,\mu_1,\mu_2) = x_1x_2+\mu_1(2-x_1-x_2)+\mu_2(x_1-x_2)$

$\frac{\partial L}{\partial x_1} = x_2-\mu_1+\mu_2 = 0$ (1)

$\frac{\partial L}{\partial x_2} = x_1-\mu_1-\mu_2 = 0$ (2)

$\mu_1(2-x_1-x_2) \pm \mu_2(x_1-x_2) = 0$ (3)

$\mu_1,\mu_2\geq 0$ (4)

$2-x_1-x_2\leq 0$ (5)

$x_1-x_2\leq 0$ (6)

(b) (1),(2) $\Rightarrow x_1=\mu_1+\mu_2$, $x_2=\mu_1-\mu_2 \Rightarrow x_1-x_2 = 2\mu_2 \overset{(6)}{\leq} 0 \overset{(4)}{\Longrightarrow}$

$\mu_2=0 \Rightarrow x_1=x_2=\mu_1 \overset{(5)}{\Longrightarrow} 2-2\mu_1\leq 0 \Rightarrow \mu_1\geq 1, \mu_1\neq 0 \overset{(3)}{\Longrightarrow}$

$2-x_1-x_2=0 \Rightarrow x_1^*=x_2^*=\mu_1^*=1$ and $\mu_2^*=0$

(c) KKT-SONC $Dg_1(x_1,x_2)=[-1,-1]$, $Dg_2(x_1,x_2)=[1,-1]$

So every point is regular!

$J(x_1^*,x_2^*)=J(1,1)=\{1,2\}$, $T(x_1^*,x_2^*)=\left\{y\in\mathbb{R}^2 \,\middle|\, \begin{matrix}[-1,-1]y=0\\ [1,-1]y=0\end{matrix}\right\}$

$\begin{cases} y_1+y_2=0\\ y_1-y_2=0\end{cases} \Rightarrow y_1=y_2=0$, $T(x_1^*,x_2^*)=\left\{\begin{bmatrix}0\\ 0\end{bmatrix}\right\}$

So KKT-SONC holds

(d) KKT-SOSC $D_x^2 L(x_1,x_2,\mu_1,\mu_2) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$$\tilde{J}(1,1;1,0) = \overbrace{\{1,2\}}^{\text{active cnsts.}} \cap \overbrace{\{1\}}^{\mu>0} = \{1\}$$

$$\tilde{T}(1,1;1,0) = \{y \mid Dg_1(x_1^*,x_2^*)y = 0\} = \{y \mid [-1,-1]y = 0\} = \left\{\begin{bmatrix} a \\ -a \end{bmatrix} \,\middle|\, a \in \mathbb{R}\right\}$$

$$y_1 + y_2 = 0$$

$$[a,-a]\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a \\ -a \end{bmatrix} = -2a^2 < 0 \text{ for } a \neq 0 \text{ so KKT-SOSC fails}$$

Example 21.12 p. 504 Quadratic Programming II

Minimize $\frac{1}{2}x^TQx$ subject to $Ax \le b$ where $Q = Q^T > 0$, $A \in \mathbb{R}^{m\times n}$, $\boxed{b \geq 0}$

KKT-FONC! $g(x) = Ax - b \le 0$

$$L(x,\mu) = \frac{1}{2}x^TQx + \mu^T(Ax-b)$$

KKT-FONC

$$\begin{cases} D_xL(x,\mu) = x^TQ + \mu^TA = 0 \ (1) \overset{\text{xx to the right}}{\Longrightarrow} x^TQx + \mu^TAx = 0 \\ \mu \geq 0 \ (2) \\ \mu^T(Ax-b) = 0 \ (3) \Longrightarrow \mu^TAx = \mu^Tb \\ Ax - b \le 0 \ (4) \end{cases}$$

$$\left.\begin{matrix} x^TQx + \mu^TAx = 0 \\ \mu^TAx = \mu^Tb \end{matrix}\right\} x^TQx + \mu^Tb = 0 \ (5)$$

$$\mu \geq 0, b \geq 0 \Rightarrow \mu^Tb \geq 0 \underset{(5)}{\Longrightarrow} x^TQx \le 0 \underset{\text{contrapos}}{\overset{Q>0}{\Longrightarrow}} x = 0$$

Remark. We can see that $x=0$ is feasible $(b \geq 0)$ with value $0 \le x^TQx \ \forall x$

so $x = 0$ is a global min

In reality this problem is convex (KKT-FONC are also sufficient)

so $x = 0$ is a local min (global)

6 Convex Programming

General form:

$$\text{(C) Minimize } f(x) \text{ subject to } x \in S,$$

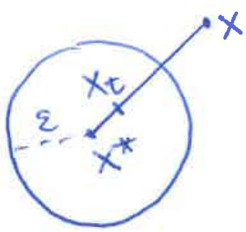
where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, that is, $\forall x,y \in \mathbb{R}^n\ \forall t \in (0,1),\ f(tx+(1-t)y) \le tf(x)+(1-t)f(y)$

and $S \subset \mathbb{R}^n$ is convex, that is, $\forall x,y \in S,\ [x,y] := \{tx+(1-t)y \mid t \in [0,1]\} \subset S$

Theorem. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex and let $S \subset \mathbb{R}^n$ be convex. Then x^* is a local minimizer of f over S iff x^* is a global minimizer of f over S.

Proof. $(\Leftarrow)$ is plain bc. global $\Rightarrow$ local

$(\Rightarrow)$ $\exists \varepsilon > 0;\ (B(x^*;\varepsilon))$, $\forall x \in S \cap B(x^*;\varepsilon)$, $f(x^*) \le f(x)$ (1)


$B(x^*;\varepsilon)$

$\forall x \in S$, $\exists t \in (0,1)$, $t < 1$ small such that

$x_t := tx + (1-t)x^* \in B(x^*;\varepsilon)$

Because $x, x^* \in S \Rightarrow x_t \in S$ since S is convex $\Big\}\ \xrightarrow{(1)}$ for $x = x_t$

$f(x^*) \le f(x_t) \overset{f \text{ convex}}{\le} t f(x) + (1-t) f(x^*)$

$\cancel{t} f(x^*) \le \cancel{t} f(x) \Rightarrow f(x^*) \le f(x),\ \forall x \in S$

So x^* is a global min of f over S QED

Corollary. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex and let $S \subset \mathbb{R}^n$ be convex. The set of all global minimizers of f over S is a convex set.

Let $\lambda = \min\limits_{x \in S} f(x)$. Then

$\underset{S}{\operatorname{argmin}} f = \{x \mid f(x) \le \lambda\}$ a level set of a convex function which is always convex!

$\Uparrow$

(the set of global mins of f over S)

6.1 Convexity Criteria

Q is positive semi def on $\Omega-\Omega$

Proposition. The quadratic form $f(x) = x^TQx$ is convex on Ω iff, for all $x, y \in \Omega$, $(x-y)^TQ(x-y) \geq 0$. Here $Q \in \mathbb{R}^{n\times n}$, $Q = Q^T$.

(monotone on Ω)

Proof. $\forall x,y\in\Omega$, $\forall t\in(0,1)$, let $x_t := tx+(1-t)y$

$$tf(x)+(1-t)f(y)-f(tx+(1-t)y) = tx^TQx+(1-t)y^TQy-(tx+(1-t)y)^TQ(tx+(1-t)y)$$

$$= tx^TQx+(1-t)y^TQy-t^2x^TQx-t(1-t)x^TQy-t(1-t)y^TQx-(1-t)^2y^TQy$$

$$= t(1-t)[x^TQx-x^TQy-y^TQx+y^TQy] = t(1-t)(x-y)^TQ(x-y)$$

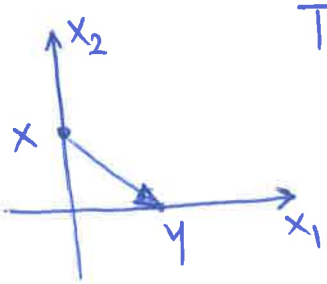
f is convex on Ω means LHS $\geq 0 \iff$ RHS $\geq 0 \iff \forall x,y\in\Omega,\ (x-y)^TQ(x-y)\geq 0$

(pick $t\neq 0,1$)

Example. Is $f(x) = x_1x_2$ convex over the first quadrant $\Omega = \{x = (x_1, x_2) \mid x_1 \geq 0,\ x_2 \geq 0\}$?

$f(x_1,x_2) = x_1\cdot x_2 = x^TQx$ for $x=\begin{bmatrix}x_1\\x_2\end{bmatrix}$ and $Q=\frac{1}{2}\begin{bmatrix}0&1\\1&0\end{bmatrix}$

Take $x=\begin{bmatrix}0\\1\end{bmatrix}$, $y=\begin{bmatrix}1\\0\end{bmatrix}\in\Omega$. Then $x-y=\begin{bmatrix}-1\\1\end{bmatrix}$



$$(x-y)^TQ(x-y) = [-1,1]\cdot\underbrace{\frac{1}{2}\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}}_{\frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix}} \cdot \frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix} = \frac{1}{2}(-2) = -1 < 0$$

So according to the previous Proposition f is not convex on Ω

Theorem. Let $f \in C^1(\Omega;\mathbb{R})$ be convex, where $\Omega \subset \mathbb{R}^n$ is open and convex. Then f is convex (on Ω) iff

$$\forall x, y \in \Omega,\ f(y) \geq f(x) + Df(x)(y-x).$$

Proof. ($\Rightarrow$) Since f is convex, by def, $\forall x,y \in \Omega$, $\forall t \in (0,1)$

$$f(ty+(1-t)x) \leq tf(y) + (1-t)f(x) \Rightarrow \frac{f(x+t(y-x)) - f(x)}{t} \leq f(y) - f(x)$$

Let $t \downarrow 0$ $\quad\downarrow$ Gâteaux deriv

$$Df(x)(y-x) \leq f(y) - f(x)$$

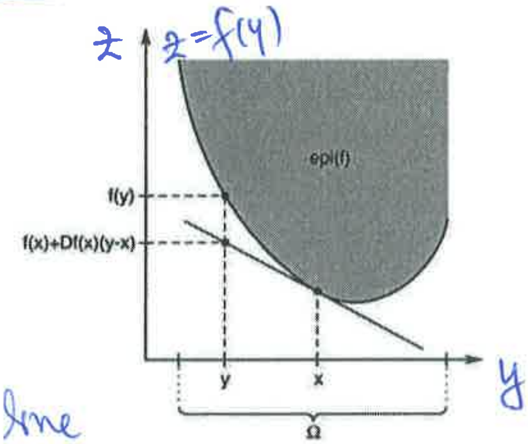
($\Leftarrow$) $\forall x,y \in \Omega$, $\forall t \in (0,1)$ let $x_t := tx + (1-t)y \in \Omega$ since Ω is convex

We have $f(x) \geq f(x_t) + Df(x_t)(x - x_t)$ / times t

$f(y) \geq f(x_t) + Df(x_t)(y - x_t)$ / times $(1-t)$

$$tf(x) + (1-t)f(y) \geq f(x_t) + \cancel{Df(x_t)(x_t - x_t)}$$

So f is convex



Geometric Interpretation. The eqn of the tangent line (linear approx) to the graph of f at $(x, f(x))$ is $z = Df(x)(y-x) + f(x)$

The tangent line lies below the graph of a (convex) function

$\boxed{\text{Extension}}$ A vector g is called a SUBGRADIENT of f at $x \in \Omega$ if

$$\forall y \in \Omega,\ f(y) \geq f(x) + g^T(y-x)$$

The set $\{g \mid g \text{ is a subgradient of } f \text{ at } x\} =: \partial f(x)$ is called the CONVEX subdifferential of f at x

The whole theory for non-smooth convex programming is based on this notion!

Theorem. Let $f \in C^2(\Omega; \mathbb{R})$ be convex, where $\Omega \subset \mathbb{R}^n$ is open and convex. Then f is convex (on Ω) iff, for every $x \in \Omega$, $D^2 f(x) \geq 0$ (is positive semidefinite). [Proof. skipped]

By definition a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is *concave* if $-f$ is convex. $\Leftrightarrow D^2 f(x) \leq 0 \ \forall x \in \Omega$

Sylvester's Criterion Let $Q = Q^T \in \mathbb{R}^{n \times n}$. Then $Q > (\geq) 0$ iff the leading principal minors of Q are positive (non-negative)

determinant of a subsquare matrix on the 1st diagonal

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ & & & \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix}$$

Example. #22.6 p.520

2. $f(x_1, x_2, x_3) = 4x_1^2 + 3x_2^2 + 5x_3^2 + 6x_1x_2 + x_1x_3 - 3x_1 - 2x_2 + 15$

$$D^2 f(x) = \begin{bmatrix} 8 & 6 & 1 \\ 6 & 6 & 0 \\ 1 & 0 & 10 \end{bmatrix}$$

$$\begin{matrix} 8 & 6 & 1 \\ 6 & 6 & 0 \end{matrix}$$

$\Delta_1 = 8 > 0$

$\Delta_2 = \begin{vmatrix} 8 & 6 \\ 6 & 6 \end{vmatrix} = 12 > 0$

$\Delta_3 = 480 + 0 + 0 - 6 - 0 - 360 = 114 > 0$

According to Sylvester's Criterion $D^2 f(x) > 0$

3. $f(x_1, x_2) = 2x_1x_2 - x_1^2 - x_2^2$ $\quad D^2 f(x) = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix}$ $\quad \Delta_1 = -2 < 0$, $\Delta_2 = 0 \leq 0$

$D^2 f(x) \leq 0 \Rightarrow -f$ is convex & f is concave

6.2 Sufficient Optimality Conditions

the domain of f

Lemma. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex such that f is $C^1(\Omega)$, where $\Omega \subset D(f)$ is an open convex set. For every convex $S \subset \Omega$ and for every $x^* \in S$

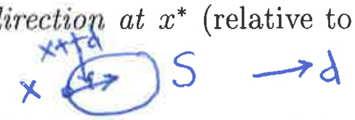
$$\forall x \in S,\ (x \neq x^*),\ Df(x^*)(x - x^*) \geq 0 \Longrightarrow x^* \in \operatorname{argmin}\{f(x) \mid x \in S\}.$$

In other words, x^* is a global minimizer of f over S.

Proof. Because f is convex on Ω, according to Thm p.62, $\forall x \in S$

$$f(x) \geq f(x^*) + \underbrace{Df(x^*)(x-x^*)}_{\geq 0} \geq f(x^*) \Rightarrow x^* \text{ is a global min of } f \text{ over } S$$ ∎

Definition. Let $S \subset \mathbb{R}^n$ and let $x^* \in S$. A vector $d \in \mathbb{R}^n$ is *a feasible direction at* x^* (relative to S) if $d \neq 0$ and there exists $t_0 > 0$ such that, for every $0 \leq t \leq t_0$, $x + td \in S$.



When S is convex and $x, y \in S$, $y \neq x$, the direction $d = y - x$ is feasible at x relative to S. Indeed

$x + td = x + t(y-x) =: x_t$

$\exists t_0 = 1,\ \forall 0 \leq t \leq 1,\ x_t = x + td = x + t(y-x) = ty + (1-t)x \in S$

since S is convex!

Theorem. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex such that f is $C^1(\Omega)$, where $\Omega \subset D(f)$ is an open convex set. Let $S \subset \Omega$ be convex. Suppose that $x^* \in S$ has the property that for every feasible direction d at x^* (relative to S), $d^T \nabla f(x^*) \geq 0$. Then x^* is a global minimizer of f over S.

Proof. $\forall x \in S$ $(x \neq x^*)$, $d = x - x^*$ is feasible at x^*. Then

$$\underbrace{Df(x^*)(x-x^*)}_{\nabla f(x^*)^T \cdot d} = d^T \cdot \nabla f(x^*) \geq 0 \text{ from our assumption}$$

According to the previous Lemma, x^* is a global min of f over S ∎

Corrolary. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex such that f is $C^1(\Omega)$, where $\Omega \subset D(f)$ is an open convex set. Suppose that $x^* \in \Omega$ has $\nabla f(x^*) = 0$. Then x^* is a global minimizer of f over Ω.

Proof. $\nabla f(x^*) = 0 \Rightarrow d^T \cdot \nabla f(x^*) = 0$ $\forall d$ a feasible direction at x^*

According to the previous Thm, x^* is a global min of f over Ω

Theorem (CNE). Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex such that f is $C^1(\Omega)$, where $\Omega \subset D(f)$ is an open convex set that contains the convex feasible set

$$S := \{x \in \mathbb{R}^n \mid h(x) = 0\},$$

where $h \in C^1(\Omega; \mathbb{R}^m)$. Suppose that there exist $x^* \in S$ and $\lambda^* \in \mathbb{R}^m$ such that

$$\text{(FONC)} \qquad Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T.$$

Then x^* is a global minimizer of f over S. (is FONC) or FONC is necessary and sufficient

Proof.

Since f is convex, $\forall x \in S$, $f(x) \geq f(x^*) + Df(x^*)(x - x^*)$ (1) (Thm. p.62)

But $Df(x^*) = -\lambda^{*T} Dh(x^*)$ so (1) becomes

$$f(x) \geq f(x^*) - \lambda^{*T} Dh(x^*)(x - x^*) \quad (2)$$

Since S is convex, $\forall x \in S, \forall t \in (0,1)$, $tx + (1-t)x^* \in S \Rightarrow$

$$h(tx + (1-t)x^*) = h(x^* + t(x - x^*)) = 0$$

We get $\dfrac{\overset{=0}{h(x^* + t(x - x^*))} - \overset{=0}{h(x^*)}}{t} = 0, \forall t \in (0,1)$. Let $t \downarrow 0$

$$\downarrow$$

$$Dh(x^*)(x - x^*) = 0 \rightarrow \lambda^{*T} Dh(x^*)(x - x^*) = 0$$

So (2) yields $f(x) \geq f(x^*)$, i.e., x^* is a global min of f over S

Theorem (CN). Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex such that f is $C^1(\Omega)$, where $\Omega \subset D(f)$ is an open convex set that contains the convex feasible set

$$S := \{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \le 0\},$$

where $h \in C^1(\Omega; \mathbb{R}^m)$, $g \in C^1(\Omega; \mathbb{R}^p)$. Suppose that there exist $x^* \in S$, $\lambda^* \in \mathbb{R}^m$, and $\mu^* \in \mathbb{R}^p$ such that

1. $\mu^* \ge 0$.
2. $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$. ie KKT-FONC are SUFFICIENT
3. $\mu^{*T} g(x^*) = 0$.

Then x^* is a global minimizer of f over S.

Proof. f convex means $\overset{\text{Thm}}{\underset{\text{p.62}}{\Longrightarrow}}$ $\forall x \in S$, $f(x) \ge f(x^*) + Df(x^*)(x-x^*)$ (1)

Relation 2. provides $Df(x^*) = -\lambda^{*T} Dh(x^*) - \mu^{*T} Dg(x^*) \overset{(1)}{\Longrightarrow}$

$\forall x \in S$, $f(x) \ge f(x^*) \underbrace{-\lambda^{*T} Dh(x^*)(x-x^*)}_{=0} - \mu^{*T} Dg(x^*)(x-x^*)$ (2)

As previously seen

(see Theorem CNE on p. 65)

We claim that $\mu^{*T} Dg(x^*)(x-x^*) \le 0$ (3)

Since S is convex $\forall x \in S$, $\forall t \in (0,1)$, $(1-t)x^* + tx \in S \Rightarrow$

$g((1-t)x^* + tx) \le 0$, $(1-t)x^* + tx = x^* + t(x-x^*)$ so

$g(x^* + t(x-x^*)) \le 0$ / times $\mu^* \ge 0$ from relation 1.

$\left.\begin{array}{l} \mu^{*T} g(x^* + t(x-x^*)) \le 0 \\ \mu^{*T} g(x^*) = 0 \text{ (relation 3.)} \end{array}\right\} \Longrightarrow \dfrac{\mu^{*T} g(x^* + t(x-x^*)) - \mu^{*T} g(x^*)}{t} \le 0$

$\Downarrow t \downarrow 0$

$\mu^{*T} Dg(x^*)(x-x^*) \le 0$ ie. (3) is proved!

Now (2) reads $f(x) \ge f(x^*)$, $\forall x \in S$ ie. x^* is a global min of f over S

■

Example. (#22.16, p. 543) Write the KKT conditions for the (Pa) **(LP)** Minimize $f(x) = c^T x$ subject to $Ax = b$, $x \geq 0$. Are the KKT conditions necessary, sufficient, or both?

$f(x) = c^T \cdot x \quad f(tx+(1-t)y) = tf(x) + (1-t)f(y), \forall t \in (0,1), \forall x,y \in \mathbb{R}^n = \Omega$

f is linear $\Rightarrow$ convex

$h(x) = Ax - b$ $\qquad g(x) = -x$ which is linear $\Rightarrow$ convex!

h affine $\Rightarrow$ convex $\qquad S = \{x \mid h(x) = 0, g(x) \leq 0\}$

$S = \underbrace{\{x \mid h(x) = 0\}}_{\text{affine set so convex too!}} \cap \underbrace{\{x \mid g(x) \leq 0\}}_{\text{level set for } g \text{ convex so it is convex}}$ convex

Another way to prove S is convex: (by definition)

$\forall x,y \in S, \forall t \in (0,1)$ I want $tx + (1-t)y \in S$

$\Downarrow$

$Ax = Ay = b,\ x,y \geq 0 \Rightarrow A(tx+(1-t)y) = tAx + (1-t)Ay = tb + (1-t)b = b$

$x \geq 0, y \geq 0 \Rightarrow tx \geq 0, (1-t)y \geq 0 \overset{+}{\Rightarrow} tx + (1-t)y \geq 0$

Hence $tx + (1-t)y \in S$

(a) $Df(x) = c^T,\ Dh(x) = A,\ Dg(x) = -I_n$

1. $\mu^* \geq 0$
2. $c^T + \lambda^{*T} A - \mu^{*T} = 0$
3. $\mu^{*T} x^* = 0$

when $x^* \in S$ means

4. $Ax^* = b$
5. $x^* \geq 0$

(b) These conditions are sufficient because every LP problem is a convex programming problem and we use theorem CN on page 66

(c) (Da) Max $b^T \cdot \lambda$ subject to $\lambda^T A \leq c^T$ $[A^T \lambda \leq c]$ (asymmetric duality) (p. 31)

(d) x^* feasible for (Pa), λ^* feasible for (Da) $\quad (c^T - \lambda^{*T} A) x^* = 0 \Rightarrow x^*$ is optimal for (Pa)

Let μ^* be such that $\mu^{*T} = c^T - \lambda^{*T} A \Leftrightarrow \mu^* = c - A^T \lambda^* \geq 0$ bc λ^* is feasible for Da

So 1. and 3. hold but also 2. from KKT-FONC $\Rightarrow x^*$ is a solution $x^*, -\lambda^*, \mu^*$

Example. #22.7 p. 526 on the board

Final Exam Practice Problems

Problem 1. Prove that if f, g are convex so is $\max\{f, g\}$. Will $\min\{f, g\}$ remain convex (or concave)?

Problem 2. Find the range of values of the parameter α for which the function

$$f(x_1, x_2, x_3) = 2x_1x_3 - x_1^2 - x_2^2 - 5x_3^2 - 2\alpha x_1x_2 - 4x_2x_3$$

is concave.

Problem 3. (a) Let f be a convex function. Prove that x^* is a global minimum point of f iff $0 \in \partial f(x^*)$.

(b) Prove that if g_1 is a subgradient of f_1 at x and g_2 is a subgradient of f_2 at x then $g_1 + g_2$ is a subgradient of $f_1 + f_2$ at x.

(c) If g is a subgradient of f at x and $\lambda > 0$ find a subgradient of λf at x.

(d) Prove that the convex subfifferential is monotone, that is, for every $g_1 \in \partial f(x_1)$, $g_2 \in \partial f(x_2)$, $(g_1 - g_2)^T(x_1 - x_2) \geq 0$.

Problem 4. Find the subgradient of the absolute value function at $x = 0$ and at $x = 1$? Extrapolate to find the subgradient of the Euclidean Norm function everywhere.

Problem 5. Consider the problem:

$$\text{Minimize } \tfrac{1}{2}\|Ax - b\|^2 \text{ subject to } x_1 + x_2 + \ldots + x_n = 1,\ x_1, x_2, \ldots, x_n \geq 0.$$

Prove that this problem is a convex optimization problem.

Problem 6. Consider the problem:

$$\text{Minimize } \|x - x_0\|^2 \text{ subject to } \|x\|^2 = 9,$$

where $x_0 = [1, \sqrt{3}]^T$. Give a geometric interpretation.

(a) Find all points satisfying the Lagrange condition for the problem.

(b) Using second-order conditions, determine whether or not each of the points in part a is a local minimizer.

(c) Is this problem convex?

Problem 7. #20.7, page 483

Problem 8. Consider the problem

$$\text{Minimize } x_1x_2 - 2x_1 \text{ subject to } x_1^2 - x_2^2 = 0,$$

a. Apply Lagrange's theorem directly to the problem to show that if a solution exists, it must be either $[1, 1]^T$ or $[-1, 1]^T$.

b. Use the second-order necessary conditions to show that $[-1, 1]^T$ cannot possibly be the solution.

c. Use the second-order sufficient conditions to show that $[1, 1]^T$ is a strict local minimizer.

Problem 9. #21.6, page 502

Problem 10. #21.23, page 507

Problem 11. #22.19, page 544

Problem 12. #22.23, page 546

Look also at other end-section problems in the book!