Chapter 3: Constrained Extrema

Math 368

May 22, 2013

Implicit Function Theorem

For scalar fn $g : \mathbb{R}^n \to \mathbb{R}$ with $\nabla g(\mathbf{x}^*) \neq \mathbf{0}$ and $g(\mathbf{x}^*) = b$,

level set $g^{-1}(b) = \{\, \mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) = b \,\}$ is locally a graph.

$\dfrac{\partial g}{\partial x_m}(\mathbf{x}^*) \neq 0$ implies

x_m is locally determined implicitly as a fn of other variables.

Tangent plane

$$0 = \frac{\partial g}{\partial x_1}(\mathbf{x}^*)(x_1 - x_1^*) + \cdots + \frac{\partial g}{\partial x_n}(\mathbf{x}^*)(x_n - x_n^*)$$

If $m = n$,

$$x_n = -x_n^* - \left(\frac{\frac{\partial g}{\partial x_1}(\mathbf{x}^*)}{\frac{\partial g}{\partial x_n}(\mathbf{x}^*)} \right)(x_1 - x_1^*) - \cdots - \left(\frac{\frac{\partial g}{\partial x_{n-1}}(\mathbf{x}^*)}{\frac{\partial g}{\partial x_n}(\mathbf{x}^*)} \right)(x_{n-1} - x_{n-1}^*)$$

linear tangent plane graph $\Rightarrow$ nonlinear level set graph

Implicit Function Theorem, Higher Dimensions

For vector fn $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^k$ C^1, k constraints and n variables,

level set for constant $\mathbf{b} \in \mathbb{R}^k$,

$$\mathbf{g}^{-1}(\mathbf{b}) = \{ \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) = b_i \text{ for } i = 1, \dots, k \}.$$

To insure a graph near $\mathbf{x}^*$ with $\mathbf{g}(\mathbf{x}^*) = \mathbf{b}$, need $\operatorname{rank}(D\mathbf{g}(\mathbf{x}^*)) = k$

i.e., gradients $\{\nabla g_i(\mathbf{x}^*)\}_{i=1}^k$ are linearly independent

Implicit Function Theorem, continued

If $\text{rank}(D\mathbf{g}(\mathbf{x}^*)) = k$, can select k $x_{m_1}, \dots, x_{m_k}$ s.t.

$$(*) \qquad \det \begin{pmatrix} \frac{\partial g_1}{\partial x_{m_1}}(\mathbf{x}^*) & \cdots & \frac{\partial g_1}{\partial x_{m_k}}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial x_{m_1}}(\mathbf{x}^*) & \cdots & \frac{\partial g_k}{\partial x_{m_k}}(\mathbf{x}^*) \end{pmatrix} \neq 0.$$

Below show (*) $\Rightarrow$ $\text{null}(D\mathbf{g}(\mathbf{x}^*))$ graph of $\mathbf{z} = (x_{m_1}, \dots, x_{m_k})$

in terms of other $n-k$ variables $\mathbf{w} = (x_{\ell_1}, \dots, x_{\ell_{n-k}})$.

Implicit Fn Thm says that for nonlinear $\mathbf{g}^{-1}(\mathbf{b})$,

$\mathbf{z} = (x_{m_1}, \dots, x_{m_k})$ are also determined implicitly
as functions of the other $n-k$ variables $\mathbf{w} = (x_{\ell_1}, \dots, x_{\ell_{n-k}})$

Theorem (**Implicit Function Theorem**)

Assume $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^k$ C^1, $\mathbf{g}(\mathbf{x}^*) = \mathbf{b}$, *&* $\operatorname{rank}(D\mathbf{g}(\mathbf{x}^*)) = k$,

then nonlinear level set $\mathbf{g}^{-1}(\mathbf{b})$ *is locally a graph near* $\mathbf{x}^*$.

If $\mathbf{z} = (x_{m_1}, \dots, x_{m_k})$ *are selected s.t.* *(*)* $\det\left(\frac{\partial g_i}{\partial x_{m_j}}(\mathbf{x}^*\right) \neq 0$,

then $x_{m_1}, \dots, x_{m_k}$ *are determined implicitly near* $\mathbf{x}^*$

as functions of the other $n-k$ *variables* $\mathbf{w} = (x_{\ell_1}, \dots, x_{\ell_{n-k}})$.

Implicitly defined $(x_{m_1}, \dots, x_{m_k}) = \mathbf{h}(x_{\ell_1}, \dots, x_{\ell_{n-k}})$ *is differentiable*

$\frac{\partial x_{m_i}}{\partial x_{\ell_j}}$ *calculated by chain rule and solving matrix equation (ImDiff)*

$$\mathbf{0} = \begin{bmatrix} \frac{\partial g_1}{\partial x_{\ell_1}} & \cdots & \frac{\partial g_1}{\partial x_{\ell_{n-k}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial x_{\ell_1}} & \cdots & \frac{\partial g_k}{\partial x_{\ell_{n-k}}} \end{bmatrix} + \begin{bmatrix} \frac{\partial g_1}{\partial x_{m_1}} & \cdots & \frac{\partial g_1}{\partial x_{m_k}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial x_{m_1}} & \cdots & \frac{\partial g_k}{\partial x_{m_k}} \end{bmatrix} \begin{bmatrix} \frac{\partial x_{m_1}}{\partial x_{\ell_1}} & \cdots & \frac{\partial x_{m_1}}{\partial x_{\ell_{n-k}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_{m_k}}{\partial x_{\ell_1}} & \cdots & \frac{\partial x_{m_k}}{\partial x_{\ell_{n-k}}} \end{bmatrix}$$

Implicit Function Theorem, continued

Hardest part of Implicit Fn Thm is that $\mathbf{h}$ exists and is differentiable.

Taking $\frac{\partial}{\partial w_j}$ of $\mathbf{b} = \mathbf{g}(\mathbf{w}, \mathbf{z})$ while considering $z_i = h_i(\mathbf{w})$ as fns of w_j,

$$\mathbf{0} = \frac{\partial \mathbf{b}}{\partial w_j} = \frac{\partial \mathbf{g}}{\partial w_j} + \sum_{i=1}^{k} \frac{\partial \mathbf{g}}{\partial z_i} \frac{\partial z_i}{\partial w_j}.$$

$$\mathbf{0} = \begin{bmatrix} \frac{\partial g_1}{\partial x_{\ell_1}} & \cdots & \frac{\partial g_1}{\partial x_{\ell_{n-k}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial x_{\ell_1}} & \cdots & \frac{\partial g_k}{\partial x_{\ell_{n-k}}} \end{bmatrix} + \begin{bmatrix} \frac{\partial g_1}{\partial x_{m_1}} & \cdots & \frac{\partial g_1}{\partial x_{m_k}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial x_{m_1}} & \cdots & \frac{\partial g_k}{\partial x_{m_k}} \end{bmatrix} \begin{bmatrix} \frac{\partial x_{m_1}}{\partial x_{\ell_1}} & \cdots & \frac{\partial x_{m_1}}{\partial x_{\ell_{n-k}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_{m_k}}{\partial x_{\ell_1}} & \cdots & \frac{\partial x_{m_k}}{\partial x_{\ell_{n-k}}} \end{bmatrix}$$

Implicit Function Theorem, continued

Group partial deriv of $\mathbf{w} = (x_{\ell_1}, \dots, x_{\ell_{n-k}})$ & $\mathbf{z} = (x_{m_1}, \dots, x_{m_k})$.

$$D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*) = \left(\frac{\partial g_i}{\partial x_{\ell_j}}(\mathbf{x}^*) \right)_{1 \le i \le k,\ 1 \le j \le n-k}$$

$$D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*) = \left(\frac{\partial g_i}{\partial x_{m_j}}(\mathbf{x}^*) \right)_{1 \le i \le k,\ n-k+1 \le j \le n}$$

$$\operatorname{rank}(D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)) = k \quad \text{or} \quad \det\left(D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)\right) \neq 0.$$

$$\mathbf{0} = D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*) + D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)\, D\mathbf{h}(\mathbf{x}^*) \qquad \text{(ImDiff)}$$

$$D\mathbf{h}(\mathbf{x}^*) = -\left(D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)\right)^{-1} D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*).$$

Write down (ImDiff) then solve. Easier than formula with inverse

$D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*)$ is matrix with determinant $\neq 0$.

$D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*)$ includes all other independent variables.

null space of $D\mathbf{g}(\mathbf{x}^*)$

$$\begin{aligned}\text{null}(D\mathbf{g}(\mathbf{x}^*)) &= \{\mathbf{v} \in \mathbb{R}^n : D\mathbf{g}(\mathbf{x}^*)\mathbf{v} = \mathbf{0}\ \} \\ &= \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \cdot \nabla g_i(\mathbf{x}^*) = 0 \ \text{ for } \ 1 \leq i \leq k\,\}\end{aligned}$$

If $\det\left(D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)\right) \neq 0$ then $\text{null}(D\mathbf{g}(\mathbf{x}^*))$ is a graph, with $\mathbf{z}$ fn of $\mathbf{w}$:

For $\mathbf{v} \in \text{null}(D\mathbf{g}(\mathbf{x}^*))$,

$$\begin{aligned}\mathbf{0} &= D\mathbf{g}(\mathbf{x}^*)\mathbf{v} = [D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*), D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)] \begin{pmatrix} \mathbf{v}_{\mathbf{w}} \\ \mathbf{v}_{\mathbf{z}} \end{pmatrix} \\ &= D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*)\mathbf{v}_{\mathbf{w}} + D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)\mathbf{v}_{\mathbf{z}}, \\ \mathbf{v}_{\mathbf{z}} &= -\left(D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)\right)^{-1} D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*)\,\mathbf{v}_{\mathbf{w}},\end{aligned}$$

If linear tangent space is graph of $\mathbf{z}$ in terms of $\mathbf{w}$, then

nonlinear level set is locally a graph of $\mathbf{z}$ in terms of $\mathbf{w}$

Implicit Function Theorem, continued

In an example, differentiate $\mathbf{C} = \mathbf{g}(\mathbf{w}, \mathbf{z})$ with respect to $\mathbf{w}$

thinking of $\mathbf{z}$ as function of $\mathbf{w}$ and

solve for unknown partial derivatives $\left(\frac{\partial z_i}{\partial w_j}\right)$ from (ImDiff)

$$\mathbf{0} = \begin{bmatrix} \frac{\partial g_1}{\partial w_1} & \cdots & \frac{\partial g_1}{\partial w_{n-k}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial w_1} & \cdots & \frac{\partial g_k}{\partial w_{n-k}} \end{bmatrix} + \begin{bmatrix} \frac{\partial g_1}{\partial z_1} & \cdots & \frac{\partial g_1}{\partial z_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial z_1} & \cdots & \frac{\partial g_k}{\partial z_k} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial w_1} & \cdots & \frac{\partial z_1}{\partial w_{n-k}} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_k}{\partial w_1} & \cdots & \frac{\partial z_k}{\partial w_{n-k}} \end{bmatrix}$$

(1) matrix $D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)$ includes all the partial derivatives with respect to the dependent variables z_j used to calculate the nonzero determinant

(2) matrix $D_{\mathbf{w}}\mathbf{g}(\mathbf{x}^*)$ includes all the partial derivatives with respect to the independent (other) variables, w_j.

Example of Changing Technology for Production

Single fixed output, Q_0.

Two inputs x and y, $Q_0 = x^a y^b$.

By changing technology, exponents a and b vary independently.

Total cost of the inputs is fixed,

$$p\,x + q\,y = 125.$$

where price are p and q

What are rates of change of inputs as functions of a and b

at $x = 5$, $y = 50$, $p = 5$, $q = 2$, $a = \frac{1}{3}$, and $b = \frac{2}{3}$?

Changing Technology, continued

Rather than use $Q_0 = x^a y^b$, take logarithm: constraints

$$g_1(x, y, a, b, p, q) = px + qy = 125 \qquad \text{and} \tag{1}$$

$$g_2(x, y, a, b, p, q) = a\ln(x) + b\ln(y) = \ln(Q_0).$$

Two eqs define x and y as functions of a, b, p, and q since

$$\det\begin{pmatrix} \dfrac{\partial g_1}{\partial x} & \dfrac{\partial g_1}{\partial y} \\ \dfrac{\partial g_2}{\partial x} & \dfrac{\partial g_2}{\partial y} \end{pmatrix} = \det\begin{bmatrix} p & q \\ \frac{a}{x} & \frac{b}{y} \end{bmatrix} = \frac{pb}{y} - \frac{qa}{x} = \frac{pbx - qay}{xy}$$

$$= \frac{5 \cdot 2 \cdot 5 - 2 \cdot 1 \cdot 50}{3 \cdot 5 \cdot 50} = -\frac{1}{15} \neq 0.$$

Changing Technology, continued

Considering x and y as functions of a, b, p, and q, and differentiating
equations with respect to (four) independent variables gives (ImDiff):

$$\mathbf{0} = \begin{bmatrix} \frac{\partial g_1}{\partial a} & \frac{\partial g_1}{\partial b} & \frac{\partial g_1}{\partial p} & \frac{\partial g_1}{\partial q} \\ \frac{\partial g_2}{\partial a} & \frac{\partial g_2}{\partial b} & \frac{\partial g_2}{\partial p} & \frac{\partial g_2}{\partial q} \end{bmatrix} + \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial a} & \frac{\partial x}{\partial b} & \frac{\partial x}{\partial p} & \frac{\partial x}{\partial q} \\ \frac{\partial y}{\partial a} & \frac{\partial y}{\partial b} & \frac{\partial y}{\partial p} & \frac{\partial y}{\partial q} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & x & y \\ \ln(x) & \ln(y) & 0 & 0 \end{bmatrix} + \begin{bmatrix} p & q \\ \frac{a}{x} & \frac{b}{y} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial a} & \frac{\partial x}{\partial b} & \frac{\partial x}{\partial p} & \frac{\partial x}{\partial q} \\ \frac{\partial y}{\partial a} & \frac{\partial y}{\partial b} & \frac{\partial y}{\partial p} & \frac{\partial y}{\partial q} \end{bmatrix}$$

2×2 matrix is one with nonzero determinant and
has partial derivatives with respect to x and y, which
is same as dependent variables (rows) of last matrix.

Changing Technology, continued

Taking first matrix across equality and inverting the 2×2 matrix,

$$\begin{bmatrix} \frac{\partial x}{\partial a} & \frac{\partial x}{\partial b} & \frac{\partial x}{\partial p} & \frac{\partial x}{\partial q} \\ \frac{\partial y}{\partial a} & \frac{\partial y}{\partial b} & \frac{\partial y}{\partial p} & \frac{\partial y}{\partial q} \end{bmatrix} = -\frac{xy}{pbx - qay} \begin{bmatrix} \frac{b}{y} & -q \\ -\frac{a}{x} & p \end{bmatrix} \begin{bmatrix} 0 & 0 & x & y \\ \ln(x) & \ln(y) & 0 & 0 \end{bmatrix}$$

$$= \frac{xy}{qay - pbx} \begin{bmatrix} -q\ln(x) & -q\ln(y) & \frac{bx}{y} & b \\ p\ln(x) & p\ln(y) & -a & -\frac{ay}{x} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{-xyq\ln(x)}{qay - pbx} & \frac{-xyq\ln(y)}{qay - pbx} & \frac{bx^2}{qay - pbx} & \frac{bxy}{qay - pbx} \\ \frac{xyp\ln(x)}{qay - pbx} & \frac{xyp\ln(y)}{qay - pbx} & \frac{-axy}{qay - pbx} & \frac{-ay^2\ln(y)}{qay - pbx} \end{bmatrix}.$$

Changing Technology, continued

At the point in question, $qay - pbx = \frac{100-50}{3} = \frac{50}{3}$ and

$$\frac{\partial x}{\partial a} = \frac{\text{-}xyq\ln(x)}{qay - pbx} = \frac{\text{-}3(5)(50)(2)\ \ln(5)}{50} = \text{-}30\ \ln(5),$$

$$\frac{\partial x}{\partial b} = \frac{\text{-}xyq\ln(y)}{qay - pbx} = \frac{\text{-}3(5)(50)(2)\ \ln(50)}{50} = \text{-}30\ \ln(50),$$

$$\frac{\partial y}{\partial a} = \frac{xyp\ln(x)}{qay - pbx} = \frac{3(5)(50)(5)\ \ln(5)}{50} = 75\ \ln(5),$$

$$\frac{\partial y}{\partial b} = \frac{xyp\ln(y)}{qay - pbx} = \frac{3(5)(50)(5)\ \ln(50)}{50} = 75\ \ln(50).$$

Outline of implicit differentiation with several constraints

- **First**, some equations are derived (or given) relating several variables.

  In the last example, there are two equations with six variables.

- **Second**, thinking of these equations as defining some variables in terms of the others, take partial derivatives to give equation (ImDiff).

  In last example, take partial derivatives with respect to x and y.

- **Finally**, solve for the matrix of partial derivatives of the dependent variables with respect to the independent variables.

Definition

C^1 constraints $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, k$ with $\mathbf{g}(\mathbf{x}) = (g_i(\mathbf{x}))$ satisfy

constraint qualification at $\mathbf{x}^*$ p.t.

$$\operatorname{rank}(D\mathbf{g}(\mathbf{x}^*)) = k.$$

i.e., gradients $\{\nabla g_i(\mathbf{x}^*)\}_{i=1}^{k}$ are linearly independent.

Method of Lagrange Multipliers

Theorem

Assume $f, g_i : \mathbb{R}^n \to \mathbb{R}$ *are* C^1 *functions for* $i = 1, \dots, k$.

Suppose that $\mathbf{x}^*$ *is a local extremizer of* f *on*

$$\mathbf{g}^{-1}(\mathbf{b}) = \{\ \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) = b_i,\ i = 1, \dots, k\ \}.$$

Then at least one of the following holds:

1. $\exists\ \boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_k^*) \in \mathbb{R}^k$ *s.t.*

   $$Df(\mathbf{x}^*) = \sum_{i=1}^k \lambda_i^*\, Dg_i(\mathbf{x}^*),$$

   i.e., $\nabla f =$ *linear combination of* ∇g_i *of constraints*

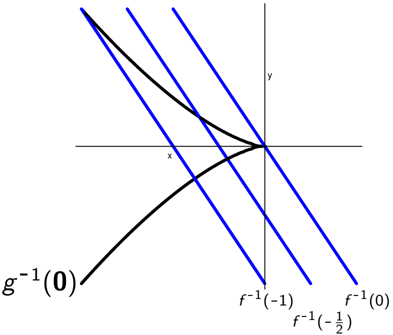
2. $\operatorname{rank}(D\mathbf{g}(\mathbf{x}^*)) < k$, *constraint qualification fails at* $\mathbf{x}^*$.

Could add condition $\operatorname{rank}(D\mathbf{g}(\mathbf{x}^*)) = k$ to assumptions.

But emphasize that max/min could come a point with lower rank.

Singular Example, $\mathbb{R}^2$

Let $g(x, y) = x^3 + y^2 = 0$ and $f(x, y) = y + 2x$.



Maximum of $f(x, y)$ is at singular point $(0, 0)$, where $\nabla g = \mathbf{0}$.

$$\nabla f(\mathbf{0}) = (2, 1)^{\mathsf{T}} \neq \mathbf{0} = \lambda \nabla g.$$

Singular Example, $\mathbb{R}^3$

$g_1(x, y, z) = x^3 + y^2 + z = 0$, $g_2(x, y, z) = z = 0$, $f(x, y, z) = y + 2x$.

Level set is that of last example in (x, y)-plane

$$\mathbf{g}^{-1}(\mathbf{0}) = \{\, (x, y, 0) : x^3 + y^2 = 0 \,\}$$

Maximum at $\mathbf{0}$.

Gradients parallel at $\mathbf{0}$:

$$\nabla g_1(x, y, z) = (3x^2,\ 2y,\ 1)^{\mathsf{T}},\ \ \nabla g_2(x, y, z) = (0,\ 0,\ 1)^{\mathsf{T}}$$

$$\nabla g_1(0, 0, 0) = (0,\ 0,\ 1)^{\mathsf{T}},\ \ \nabla g_2(0, 0, 0) = (0,\ 0,\ 1)^{\mathsf{T}}$$

$$\text{rank}(D\mathbf{g}(\mathbf{0})) = \text{rank} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = 1.$$

$$\nabla f(\mathbf{0}) = (2, 1, 0)^{\mathsf{T}} \neq \lambda_1 \nabla g_1(\mathbf{0}) + \lambda_2 \nabla g_2(\mathbf{0}) = (0, 0, \lambda_1 + \lambda_2)^{\mathsf{T}}.$$

Tangent Space to Level Set

For $\mathbf{g}(\mathbf{x}^*) = \mathbf{b}$, denote set of tangent vectors at $\mathbf{x}^*$ to $\mathbf{g}^{-1}(\mathbf{b})$ by

$$\mathbf{T}(\mathbf{x}^*) = \Big\{ \mathbf{v} = \mathbf{r}'(0) : \mathbf{r}(t) \text{ is a } C^1 \text{ curve with } \mathbf{r}(0) = \mathbf{x}^*, \; \mathbf{g}(\mathbf{r}(t)) = \mathbf{b} \text{ for all small } t \Big\}.$$

Called **tangent space of $\mathbf{g}^{-1}(\mathbf{b})$ at $\mathbf{x}^*$**

$$\mathrm{null}(D\mathbf{g}(\mathbf{x}^*)) = \{ \mathbf{v} : \mathbf{v} \cdot \nabla g_j(\mathbf{x}^*) = 0 \text{ for } j = 1, \dots, k \}.$$

Proposition (3.8)

Assume $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^k$ *is* C^1, $\mathbf{g}(\mathbf{x}^*) = \mathbf{b}$, *and* $\mathrm{rank}(D\mathbf{g}(\mathbf{x}^*)) = k$.

Then $\mathbf{T}(\mathbf{x}^*) = \mathrm{null}(D\mathbf{g}(\mathbf{x}^*))$, *set of tangent vectors is* $\mathrm{null}(D\mathbf{g}(\mathbf{x}^*))$

Most calculus books just state that set of tangent vectors are

all vectors perpendicular to gradients of constraints, $= \mathrm{null}(D\mathbf{g}(\mathbf{x}^*))$.

Proof:

For any curve $\mathbf{r}(t)$ in $\mathbf{g}^{-1}(\mathbf{b})$ with $\mathbf{r}(0) = \mathbf{x}^*$,

$$\mathbf{0} = \frac{d}{dt}\,\mathbf{b} = \frac{d}{dt}\mathbf{g}(\mathbf{r}(t))\Big|_{t=0} = D\mathbf{g}(\mathbf{r}(0))\mathbf{r}'(0) = D\mathbf{g}(\mathbf{x}^*)\mathbf{r}'(0),$$

so $\mathbf{T}(\mathbf{x}^*) \subset \text{null}(D\mathbf{g}(\mathbf{x}^*))$.

Proof of $\text{null}(D\mathbf{g}(\mathbf{x}^*)) \subset \mathbf{T}(\mathbf{x}^*)$ uses Implicit Function Theorem:

Assume variables have been ordered $\mathbf{x} = (\mathbf{w}, \mathbf{z})$ so that

$\det(D_{\mathbf{z}}\mathbf{g}(\mathbf{x}^*)) \neq 0$.

Then $\mathbf{g}^{-1}(\mathbf{b})$ is locally a graph $\mathbf{z} = \mathbf{h}(\mathbf{w})$.

Proof, continued

For $\mathbf{v} = (\mathbf{v_w}, \mathbf{v_z})^T \in \text{null}(D\mathbf{g}(\mathbf{x}^*))$, $\mathbf{w}(t) = \mathbf{w}^* + t\,\mathbf{v_w}$ is line in $\mathbf{w}$-space

$\mathbf{r}(t) = \begin{pmatrix} \mathbf{w}(t) \\ \mathbf{h}(\mathbf{w}(t)) \end{pmatrix}$ is a curve in level set.

$\mathbf{r}(0) = \mathbf{x}^*$, $\mathbf{r}'(0) = \begin{pmatrix} \mathbf{v_w} \\ D\mathbf{h}(\mathbf{w}^*)\,\mathbf{v_w} \end{pmatrix} \in \mathbf{T}(\mathbf{x}^*)$, and

$$0 = \frac{d}{dt}\, g(\mathbf{r}(t))\Big|_{t=0} = D\mathbf{g}(\mathbf{x}^*)\mathbf{r}'(0)$$

$\mathbf{r}'(0) \in \text{null}(D\mathbf{g}(\mathbf{x}^*))$ and has same $\mathbf{w}$-components as $\mathbf{v}$.

$\text{null}(D\mathbf{g}(\mathbf{x}^*))$ is a graph over the $\mathbf{w}$-coordinates, so

$$\mathbf{v} = \mathbf{r}'(0) \in \mathbf{T}(\mathbf{x}^*).$$

$\text{null}(D\mathbf{g}(\mathbf{x}^*)) \subset \mathbf{T}(\mathbf{x}^*)$.

Combining, $\text{null}(D\mathbf{g}(\mathbf{x}^*)) = \mathbf{T}(\mathbf{x}^*)$. QED

Proof of Lagrange Theorem.

Assume $\mathbf{x}^*$ is a local extremizer of f on $\mathbf{g}^{-1}(\mathbf{b})$, $\text{rank}(Dg(\mathbf{x}^*)) = k$.

For any curve $\mathbf{r}(t)$ in $\mathbf{g}^{-1}(\mathbf{b})$ with $\mathbf{r}(0) = \mathbf{x}^*$ and $\mathbf{r}'(0) = \mathbf{v}$:

$$0 = \frac{d}{dt} f(\mathbf{r}(t))\bigg|_{t=0} = Df(\mathbf{x}^*)\,\mathbf{v}$$

i.e., $Df(\mathbf{x}^*)\mathbf{v} = 0$ for all $\mathbf{v} \in \mathbf{T}(\mathbf{x}^*) = \text{null}(D\mathbf{g}(\mathbf{x}^*))$

$$\text{null}(D\mathbf{g}(\mathbf{x}^*)) = \text{null}\begin{pmatrix} D\mathbf{g}(\mathbf{x}^*) \\ Df(\mathbf{x}^*) \end{pmatrix}$$

rank $=$ #columns $-$ dim(null)

$$\text{rank}\begin{pmatrix} D\mathbf{g}(\mathbf{x}^*) \\ Df(\mathbf{x}^*) \end{pmatrix} = \text{rank}((D\mathbf{g}(\mathbf{x}^*)) = k$$

Last row, $Df(\mathbf{x}^*)$, is a linear combination of first k rows,

$$Df(\mathbf{x}^*) = \textstyle\sum_{i=1}^{k} \lambda_i^* \, D(g_i)(\mathbf{x}^*).$$ □

Example 1

Find highest point satisfying $x + y + z = 12$ and $z = x^2 + y^2$.

$$\begin{aligned} \text{Maximize:} \quad & f(x, y, z) = z \\ \text{Subject to:} \quad & g(x, y, z) = x + y + z = 12 \text{ and} \\ & h(x, y, z) = x^2 + y^2 - z = 0. \end{aligned}$$

Constraint qualification: If $\nabla g = (1, 1, 1)^{\mathsf{T}} = s\,\nabla h = s\,(2x, 2y, \text{-}1)^{\mathsf{T}}$,

$s = \text{-}1,\ x = y = \text{-}\frac{1}{2}$.

To be on level set

$$z = x^2 + y^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$g\left(\text{-}\tfrac{1}{2}, \text{-}\tfrac{1}{2}, \tfrac{1}{2}\right) = \text{-}\tfrac{1}{2} \neq 12$$

No points on level set where constraint qualification fails.

Example 1, continued

Maximize: $f(x, y, z) = z$

Subject to: $g(x, y, z) = x + y + z = 12$ and

$h(x, y, z) = x^2 + y^2 - z = 0.$

First order conditions:

$$f_x = \lambda g_x + \mu h_x, \qquad 0 = \lambda + \mu 2x,$$

$$f_y = \lambda g_y + \mu h_y, \qquad 0 = \lambda + \mu 2y,$$

$$f_z = \lambda g_z + \mu h_z, \qquad 1 = \lambda - \mu.$$

$\lambda = 1 + \mu$, eliminate this variable:

$$0 = 1 + \mu + 2\mu x,$$

$$0 = 1 + \mu + 2\mu y.$$

Subtracting 2nd from 1st, $0 = 2\mu(x - y)$, so $\mu = 0$ or $x = y$.

Case $\mu = 0$: $0 = 1 + \mu + 2\mu x = 1$, contradiction.

Example 1, continued

Case: $y = x$.

$z = x^2 + y^2 = 2x^2$ and $12 = 2x + z = 2x + 2x^2$, so

$0 = x^2 + x - 6 = (x+3)(x-2)$, and $x = 2$, -3.

$x = y = 2$: $z = 2x^2 = 8$, $0 = 1 + \mu(1+2x) = 1 + 5\,\mu$,

$\mu = -\frac{1}{5}$, and $\lambda = 1 + \mu = \frac{4}{5}$.

$x = y = -3$: $z = 2x^2 = 18$, $0 = 1 + \mu(1+2x) = 1 - 5\mu$, $\mu = \frac{1}{5}$,

and $\lambda = 1 + \mu = \frac{6}{5}$.

$(\lambda^*, \mu^*, x^*, y^*, z^*) = \left(\frac{4}{5}, -\frac{1}{5}, 2, 2, 8, \right)$ and $\left(\frac{6}{5}, \frac{1}{5}, -3, -3, 18\right)$.

$f(2,2,8) = 8$ and $f(-3,-3,18) = 18$.

Constraint set is compact so extrema exist.

Maximizer $(-3, -3, 18)$.
Minimizer $(2, 2, 8)$

End of Ex

Lagrangian

Lagrangian: mnemonic device not a proof of conditions.

$$L(\boldsymbol{\lambda}, \mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i \left(b_i - g_i(\mathbf{x})\right).$$

$\mathbf{x}^*$ satisfies first order Lagrange multiplier conditions with multipliers $\boldsymbol{\lambda}^*$ iff $(\boldsymbol{\lambda}^*, \mathbf{x}^*)$ is a critical point of L with respect to all its variables,

$$\frac{\partial L}{\partial \lambda_i}(\boldsymbol{\lambda}^*, \mathbf{x}^*) = b_i - g_i(\mathbf{x}^*) = 0 \qquad \text{for } 1 \leq i \leq k \text{ and}$$

$$D_{\mathbf{x}} L(\boldsymbol{\lambda}^*, \mathbf{x}^*) = Df(\mathbf{x}^*) - \sum_{i=1}^{k} \lambda_i^* \, Dg_i(\mathbf{x}^*) = \mathbf{0}.$$

To insure that constraint qualification does not fail, need

$$k = \operatorname{rank}\left(Dg(\mathbf{x}^*)\right) = \operatorname{rank}\left(L_{\lambda_i, x_j}(\boldsymbol{\lambda}^*, \mathbf{x}^*)\right)$$

Interpretation of Lagrange Multipliers

How does $\max\{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{g}^{-1}(\mathbf{b}) \}$ **changes with changes in** b_ℓ**?**

Theorem

Assume that $f, g_i : \mathbb{R}^n \to \mathbb{R}$ *are* C^1 *with* $1 \leq i \leq k < n$.

For $\mathbf{b} \in \mathbb{R}^k$, *let* $\mathbf{x}^*(\mathbf{b})$ *be sol'n of* 1^{st}*-order Lagrange multiplier conditions*

for "nondegenerate" extremum of f *on* $\mathbf{g}^{-1}(\mathbf{b})$

with multipliers $\lambda_1^*(\mathbf{b}), \dots, \lambda_k^*(\mathbf{b})$ *and* $\text{rank}(Dg(\mathbf{x}^*(\mathbf{b}))) = k$.

Then,

$$\lambda_i^*(\mathbf{b}) = \frac{\partial}{\partial b_i} f(\mathbf{x}^*(\mathbf{b})).$$

Marginal value of i^{th}-resource equals Lagrange multiplier

Like discussion for duality in linear programming

Proof of Interpretation:

Use Lagrangian as a function of $\mathbf{b}$ as well as $\mathbf{x}$ and $\boldsymbol{\lambda}$,

$$L(\boldsymbol{\lambda}, \mathbf{x}, \mathbf{b}) = f(\mathbf{x}) + \sum_{j=1}^{k} \lambda_j \left(b_j - g_j(\mathbf{x})\right).$$

For $\mathbf{b}$ fixed, $(\boldsymbol{\lambda}^*, \mathbf{x}^*) = (\boldsymbol{\lambda}^*(\mathbf{b}), \mathbf{x}^*(\mathbf{b}))$ satisfy

$$\mathbf{0} = \begin{pmatrix} L_{\lambda_i} \\ L_{x_i} \end{pmatrix} = \begin{pmatrix} b_i - g_i(\mathbf{x}) \\ Df(\mathbf{x})^{\mathsf{T}} - \sum_j \lambda_j \, Dg_j(\mathbf{x})^{\mathsf{T}} \end{pmatrix} = \mathbf{G}(\boldsymbol{\lambda}, \mathbf{x}, \mathbf{b})$$

$$(\, D_{\boldsymbol{\lambda}}\mathbf{G}, D_{\mathbf{x}}\mathbf{G} \,) = \begin{bmatrix} \mathbf{0}_k & \text{-}Dg \\ \text{-}Dg^{\mathsf{T}} & D_{\mathbf{x}}^2 L \end{bmatrix}, \qquad \text{“bordered Hessian”}.$$

If $\mathbf{x}^*(\mathbf{b})$ is nondegenerate extremizer on $\mathbf{g}^{-1}(\mathbf{b})$, then

$$\det(D_{\boldsymbol{\lambda}}\mathbf{G}, D_{\mathbf{x}}\mathbf{G})(\boldsymbol{\lambda}^*(\mathbf{b}), \mathbf{x}^*(\mathbf{b}), \mathbf{b}) = \det \begin{bmatrix} \mathbf{0}_k & \text{-}Dg \\ \text{-}Dg^{\mathsf{T}} & D_{\mathbf{x}}^2 L \end{bmatrix} \neq 0.$$

See Addendum 3.5 of online class book

Therefore, $\mathbf{x}^*(\mathbf{b})$ and $\boldsymbol{\lambda}^*(\mathbf{b})$ are differentiable functions of $\mathbf{b}$.

Proof of Interpretation, continued

$L(\boldsymbol{\lambda}, \mathbf{x}, \mathbf{b}) = f(\mathbf{x}) + \sum_{j=1}^{k} \lambda_j \left(b_j - g_j(\mathbf{x})\right).$

$$D_{\boldsymbol{\lambda}} L(\mathbf{x}^*) = (b_1 - g_1(\mathbf{x}^*), \dots, b_k - g_k(\mathbf{x}^*)) = \mathbf{0}$$

$$D_{\mathbf{x}} L(\mathbf{x}^*) = Df(\mathbf{x}^*) - \sum_{j=1}^{k} \lambda_j^* Dg_j(\mathbf{x}^*) = \mathbf{0}$$

$$\frac{\partial L}{\partial b_i}(\mathbf{x}^*(\mathbf{b}), \boldsymbol{\lambda}^*(\mathbf{b}), \mathbf{b}) = \lambda_i^*(\mathbf{b}).$$

$f(\mathbf{x}^*(\mathbf{b})) = L(\boldsymbol{\lambda}^*(\mathbf{b}), \mathbf{x}^*(\mathbf{b}), \mathbf{b})$

$$\begin{aligned} \frac{\partial}{\partial b_i} f(\mathbf{x}^*(\mathbf{b})) &= D_{\boldsymbol{\lambda}} L \, \frac{\partial}{\partial b_i} \boldsymbol{\lambda}^*(\mathbf{b}) + D_{\mathbf{x}} L \, \frac{\partial}{\partial b_i} \mathbf{x}^*(\mathbf{b}) + \frac{\partial L}{\partial b_i}(\mathbf{x}^*(\mathbf{b}), \boldsymbol{\lambda}^*(\mathbf{b}), \mathbf{b}) \\ &= \lambda_i^*(\mathbf{b}). \end{aligned}$$

QED

3.3 Inequality Constraints, Necessary Conditions

Maximize $f(\mathbf{x})$ on $\mathscr{F}_{\mathbf{g},\mathbf{b}} = \{\, \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq b_i \quad i = 1, \dots, m \,\}$

$g_i(\mathbf{x}) \leq b_i$ is **slack at $\mathbf{x} = \mathbf{p} \in \mathscr{F}$** p.t. $g_i(\mathbf{p}) < b_i$.

$g_i(\mathbf{x}) \leq b_i$ is **effective or tight at $\mathbf{x} = \mathbf{p} \in \mathscr{F}$** p.t. $g_i(\mathbf{p}) = b_i$.

($=$ in constraint, $\mathbf{p}$ is on the boundary of $\mathscr{F}$.)

$\mathbf{E}(\mathbf{p}) = \{\, i : g_i(\mathbf{p}) = b_i \,\}$ be set of tight constraints at $\mathbf{p}$,

$|\mathbf{E}(\mathbf{p})|$ be cardinality of $\mathbf{E}(\mathbf{p})$, and

$\mathbf{g}_{\mathbf{E}(\mathbf{p})}(\mathbf{x}) = (\, g_i(\mathbf{x}) \,)_{i \in \mathbf{E}(\mathbf{p})}$.

$\mathbf{g}(\mathbf{x})$ satisfies **constraint qualification at $\mathbf{p}$** p.t.

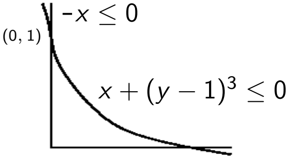
$$\operatorname{rank}(D\mathbf{g}_{\mathbf{E}(\mathbf{p})}(\mathbf{p})) = |\mathbf{E}(\mathbf{p})|,$$

gradients of tight constraints are linearly independent.

Constraint Qualification Fails

Consider the constraints

$$g_1(x,y) = x + (y-1)^3 \le 0 \qquad (\ x \le -(y-1)^3\ )$$
$$g_2(x,y) = -x \le 0$$
$$g_3(x,y) = -y \le 0$$



At $(0,1)$, $\mathbf{E}(0,1) = \{1,2\}$, $\qquad g_{\mathbf{E}(0,1)}(x,y) = (x + (y-1)^3, -x)^{\mathsf{T}}$,

$$\operatorname{rank}(D\mathbf{g}_{\mathbf{E}(0,1)}(0,1)) = \operatorname{rank}\begin{bmatrix} 1 & 3(y-1)^2 \\ -1 & 0 \end{bmatrix}_{y=1} = \operatorname{rank}\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix} = 1$$
$$< 2 = |\mathbf{E}(0,1)|$$

Constraint qualification fails at $(0,1)$.

Theorem (3.12 KKT Necessary Condtions for Extrema w/ Ineq)

Suppose that $f, g_i : \mathbf{U} \to \mathbb{R}$ *are* C^1 *functions for* $1 \le i \le m$ *where* $\mathbf{U} \subset \mathbb{R}^n$ *is open,*

$$\mathscr{F} = \{\, \mathbf{x} \in \mathbf{U} : g_i(\mathbf{x}) \le b_i \;\; \text{for} \;\; i = 1, \dots, m \,\}.$$

If f *attains a local extrema at* $\mathbf{x}^*$ *on* $\mathscr{F}$, *then either*

(**a**) *the constraint qualification fails at* $\mathbf{x}^*$,

$$\operatorname{rank}(D\mathbf{g}_{\mathbf{E}(\mathbf{x}^*)}(\mathbf{x}^*)) < |\mathbf{E}(\mathbf{x}^*)|, \text{ or}$$

(**b**) *there exist* $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)$ *such that KKT-1,2 hold:*

KKT-1. $Df(\mathbf{x}^*) = \sum_{i=1}^m \lambda_i^* \, Dg_i(\mathbf{x}^*)$.

KKT-2. $\lambda_i^* \, (b_i - g_i(\mathbf{x}^*)) = 0$ *for* $1 \le i \le m$
(*so* $\lambda_i^* = 0$ *for* $i \notin \mathbf{E}(\mathbf{x}^*)$).

KKT-3. *If* $\mathbf{x}^*$ *is a local maximum, then* $\lambda_i^* \ge 0$ *for* $1 \le i \le m$.

KKT-3′. *If* $\mathbf{x}^*$ *is a local minimum, then* $\lambda_i^* \le 0$ *for* $1 \le i \le m$.

Necessary Cond for Extrema, cont.

Homewk probl with max at point where constraint qualification fails.

Call **KKT-1,2,3** **first order KKT conditions**. More direct to use, and equivalent to derivatives of Lagrangian.

KKT-1 $\nabla f(\mathbf{x}^*)$ perpendicular to tangent space to $\mathbf{g}^{-1}(\mathbf{b})$

KKT-2 $\lambda_i^* (b_i - g_i(\mathbf{x}^*)) = 0$ is called **complementary slackness** because both $g_i(\mathbf{x}^*) \leq b_i$ & $\lambda_i^* \geq 0$ can't be slack

All $\lambda_i^* \geq 0$ at max $\mathbf{x}^*$ means $\nabla f(\mathbf{x}^*)$ points out of feasible set

Inequalities $g_i(\mathbf{x}) \leq b_i$ are resource type and signs like Max Lin Prog

All $\lambda_i^* \leq 0$ at min $\mathbf{x}^*$ means $\nabla f(\mathbf{x}^*)$ points into feasible set

Maximizes $-f(\mathbf{x})$ & $-\nabla f(\mathbf{x}^*) = \sum_{i=1}^{m} (-\lambda_i^*) \nabla g_i(\mathbf{x}^*)$

with $-\lambda_i^* \geq 0$, so signs compatible with Min Lin Prog

Proof

Assume constraint qualification holds at maximizer $\mathbf{x}^*$

Rearrange indices of g_j so that $\mathbf{E}(\mathbf{x}^*) = \{1, \ldots, k\}$,

$$g_i(\mathbf{x}^*) = b_i \text{ for } 1 \le i \le k \text{ and } g_i(\mathbf{x}^*) < b_i \text{ for } k+1 \le i \le m.$$

Rearrange indices of x_j so that $\det \left(\frac{\partial g_i}{\partial x_j}(\mathbf{x}^*) \right)_{1 \le i,j \le k} \neq 0.$

Set $\lambda_i^* = 0$ for $i \notin \mathbf{E}(\mathbf{x}^*)$, i.e., $\lambda_i^* = 0$ for $k+1 \le i \le m$.

(Drop these ineffective constraints in the argument.)

f attains a maximum at $\mathbf{x}^*$ on

$\{\mathbf{x} : g_i(\mathbf{x}) = b_i \text{ for } i \in \mathbf{E}(\mathbf{x}^*)\}$, so by Lagrange Mult Thm,

$\exists\ \lambda_i^*$ for $1 \le i \le k$ so that $\quad Df(\mathbf{x}^*) = \sum_{1 \le i \le k} \lambda_i^* \, Dg_i(\mathbf{x}^*)$.

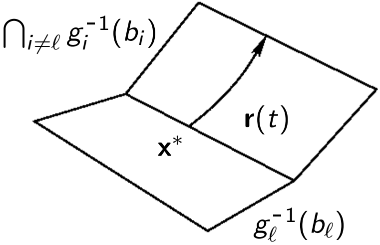
Since $\lambda_i^* = 0$ for $k+1 \le i \le m$,

$Df(\mathbf{x}^*) = \sum_{1 \le i \le m} \lambda_i^* \, Dg_i(\mathbf{x}^*) \qquad$ KKT-1 hold

Proof, continued

$\lambda_i^* = 0$ for $i \notin \mathbf{E}(\mathbf{x}^*)$ & $b_i - g_i(\mathbf{x}^*) = 0$ for $i \in \mathbf{E}(\mathbf{x}^*)$ so KKT-2 holds

Why are $\lambda_\ell^* \geq 0$ for $\ell \in \mathbf{E}(\mathbf{x}^*)$? KKT-3 for maximizer



Want a curve $\mathbf{r}(t)$ in $\mathscr{F}$ such that $g_\ell(\mathbf{r}(t)) < b_\ell$ for $t > 0$,

$g_i(\mathbf{r}(t)) = b_i$ for $i \neq \ell$ & $1 \leq i \leq k$,

$r_i(t) = x_i^*$ for $k+1 \leq i \leq n$. $\quad \det\left(\dfrac{\partial g_i}{\partial x_j}(\mathbf{x}^*)\right)_{1 \leq i,j \leq k} \neq 0.$

Let $\delta_{i\ell} = 0$ if $i \neq \ell$, & $\delta_{\ell\ell} = 1$.

Proof, continued

Apply Implicit Function Theorem to

$$G_i(\mathbf{x}, t) = g_i(\mathbf{x}) - b_i + \delta_{i\ell}\, t \quad \text{for } 1 \leq i \leq k = |\mathbf{E}(\mathbf{x}^*)|$$

$$G_i(\mathbf{x}, t) = x_i - x_i^* \text{ for } k+1 \leq i \leq n. \qquad \mathbf{G}(\mathbf{x}^*, 0) = \mathbf{0}.$$

$$\det(D_{\mathbf{x}}\mathbf{G}(\mathbf{x}^*, 0)) = \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_k} & \frac{\partial g_1}{\partial x_{k+1}} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial x_1} & \cdots & \frac{\partial g_k}{\partial x_k} & \frac{\partial g_k}{\partial x_{k+1}} & \cdots & \frac{\partial g_k}{\partial x_n} \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

$$= \det \left(\frac{\partial g_i}{\partial x_j}(\mathbf{x}^*) \right)_{1 \leq i,j \leq k} \neq 0.$$

Proof, continued

By Implicit Function theorem, there exists $\mathbf{x} = \mathbf{r}(t)$ such that

$\mathbf{r}(0) = \mathbf{x}^*$ and $\mathbf{G}(\mathbf{r}(t), t) \equiv \mathbf{0}$:

$g_i(\mathbf{r}(t)) = b_i - \delta_{i\ell}\, t$ for $1 \leq i \leq k$, $\quad r_i(t) = x_i^*$ for $k+1 \leq i \leq n$.

$$Dg_i(\mathbf{x}^*)\,\mathbf{r}'(0) = \tfrac{d}{dt}\, g_i \circ \mathbf{r}(t)\Big|_{t=0} = -\delta_{i\ell} \text{ for } 1 \leq i \leq k.$$

$f(\mathbf{x}^*) \geq f(\mathbf{r}(t))$ for $t \geq 0$, so

$$\begin{aligned} 0 \geq \frac{d}{dt} f \circ \mathbf{r}(t)\Big|_{t=0} &= Df(\mathbf{x}^*)\,\mathbf{r}'(0) \\ &= \textstyle\sum_{1 \leq i \leq k} \lambda_i^*\, Dg_i(\mathbf{x}^*)\,\mathbf{r}'(0) \\ &= \textstyle\sum_{1 \leq i \leq k} -\lambda_i^*\, \delta_{i\ell} \\ &= -\lambda_\ell^*. \end{aligned}$$

$\lambda_\ell^* \geq 0$. **QED**

Steps to use KKT

1. Verify that a maximum (resp. minimum) exists

by showing either that the feasible set is compact

or that $f(\mathbf{x})$ takes on smaller values (resp. larger values) near ∞.

2. Find all possible extremizers:

(i) Find all points on $\partial(\mathscr{F})$ where constraint qualification fails;

(ii) find all points that satisfy KKT-1,2,3 (resp. KKT-1,2,3$'$).

3. Compare $f(\mathbf{x})$ at all points found in 2(i) and 2(ii).

Example 2

Let $f(x, y) = x^2 - y$ and $g(x, y) = x^2 + y^2 \leq 1$.

Constraint set is compact so max & min exist

Derivative of constraint is $Dg(x, y) = (2x, 2y)$,

which has rank one at all points on boundary, where $g(x, y) = 1$

(At least one variable is nonzero at each of points.)

Constraint qualification is satisfied at all points in $\mathscr{F}$.

KKT-1,2 are

$$0 = f_x - \lambda\, g_x = 2x - \lambda\, 2x = 2x\,(1 - \lambda),$$
$$0 = f_y - \lambda\, g_y = -1 - \lambda\, 2y,$$
$$0 = \lambda\,(1 - x^2 - y^2).$$

From 1st equation, $x = 0$ or $\lambda = 1$.

Example 2, continued

Case (i): $\lambda = 1 > 0$. Left with equations

$$1 = -2y,$$
$$1 = x^2 + y^2.$$

$$y = \text{-}\tfrac{1}{2}, \quad x^2 = 1 - \tfrac{1}{4} = \tfrac{3}{4}, \quad x = \frac{\pm\sqrt{3}}{2}.$$

Case (ii): $x = 0$.

$$1 = -\lambda\, 2y$$
$$0 = \lambda\,(1 - y^2).$$

$\lambda \neq 0$ from 1st equation.

$y = \pm 1$ from 2nd equation

If $y = 1$: $2\lambda = -1 \;\; \lambda = \text{-}\frac{1}{2} < 0$.

If $y = \text{-}1$: $2\lambda = 1 \;\; \lambda = \frac{1}{2} > 0$.

Example 2, continued

$\lambda > 0$ for $\left(\frac{\pm\sqrt{3}}{2}, -\frac{1}{2}\right)$ & $(0, -1)$,

$$f\left(\frac{\pm\sqrt{3}}{2}, -\frac{1}{2}\right) = \frac{3}{4} + \frac{1}{2} = \frac{5}{4} \text{ and}$$

$$f(0, -1) = 1.$$

maximum is $\frac{5}{4}$, attained at $\left(\frac{\pm\sqrt{3}}{2}, -\frac{1}{2}\right)$.

$\lambda < 0$ for $(0, 1)$ $\quad f(0, 1) = -1$ minimum

$(0, -1)$ saddle: $\lambda > 0$ so decreases into $\mathscr{F}$,

but local min within boundary, so not a local maximum.

End of Example

Example 3

Maximize $f(x, y, z) = x^2 + 2y^2 + 3z^2$,

Subject to:

$$1 = x + y + z = g_0(x, y, z),$$

$$0 \geq -x = g_1(x, y, z),$$

$$0 \geq -y = g_2(x, y, z), \text{ and}$$

$$0 \geq -z = g_3(x, y, z).$$

Feasible set $\mathscr{F}$ compact, so max exists

Check constraint qualification at all points of $\mathscr{F}$:

$g_0(x, y, z) = 1$ and $g_i(x, y, z) < 0$ for $i = 1, 2, 3$

$$\text{rank}\left(D\mathbf{g_E}(x, y, z)\right) = \text{rank}\left(Dg_0(x, y, z)\right) = \text{rank}\left(\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\right) = 1,$$

$$\text{rank}\left(D\mathbf{g_E}(0, y, z)\right) = \text{rank}\left(D(g_0, g_1)^T\right) = \text{rank}\left(\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \end{bmatrix}\right) = 2$$

Example 3, constrain qualification continued

$$\text{rank}\left(D(g_0, g_2)^T\right)(x, 0, z) = \text{rank}\left(\begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \end{bmatrix}\right) = 2$$

$$\text{rank}\left(D(g_0, g_3)^T\right)(x, y, 0) = \text{rank}\left(\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}\right) = 2$$

At vertices where three constraints are tight,

$$\text{rank}\left(D(g_0, g_2, g_3)^T(1, 0, 0)\right) = \text{rank}\left(\begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\right) = 3$$

Similarly

$$\text{rank}\left(D\mathbf{g_E}(0, 1, 0)\right) = \text{rank}\left(D(g_0, g_1, g_3)^T(0, 1, 0)\right) = 3$$

$$\text{rank}\left(D\mathbf{g_E}(0, 0, 1)\right) = \text{rank}\left(D(g_0, g_1, g_2)^T(0, 0, 1)\right) = 3$$

All ranks are as large as possible,

so constraint qualification is satisfied on feasible set.

Example 3, continued

Maximize $f(x, y, z) = x^2 + 2y^2 + 3z^2$, for $1 = x + y + z = g_0(x, y, z)$,

$$0 \geq \text{-}x = g_1(x, y, z), \;\; 0 \geq \text{-}y = g_2(x, y, z), \;\; 0 \geq \text{-}z = g_3(x, y, z).$$

KKT-1,2 are

$$0 = f_x - \lambda_0\, g_{1x} - \lambda_1\, g_{1x} - \lambda_2\, g_{2x} - \lambda_3\, g_{3x} = 2x - \lambda_0 + \lambda_1$$

$$0 = f_y - \lambda_0\, g_{1y} - \lambda_1\, g_{1y} - \lambda_2\, g_{2y} - \lambda_3\, g_{3y} = 4y - \lambda_0 + \lambda_2$$

$$0 = f_z - \lambda_0\, g_{1z} - \lambda_1\, g_{1z} - \lambda_2\, g_{2z} - \lambda_3\, g_{3z} = 6z - \lambda_0 + \lambda_3$$

$$1 = x + y + z,$$

$$0 = \lambda_1\, x, \quad 0 = \lambda_2\, y, \quad 0 = \lambda_3\, z.$$

Because 0^{th}-equation involves an equality, λ_0 can have any sign.

For $1 \leq i \leq 3$, need $\lambda_i \geq 0$.

Example 3, continued

$$\lambda_0 = 2x + \lambda_1 = 4y + \lambda_2 = 6z + \lambda_3 \qquad \text{(eliminate } \lambda_0\text{)}$$

Case 1: Point with $x > 0$, $y > 0$, and $z > 0$. $\lambda_i = 0$ for $1 \leq i \leq 3$.

$$\lambda_0 = 2x = 4y = 6z, \qquad y = \tfrac{x}{2} \qquad z = \tfrac{x}{3}$$

Substituting into g_0, $1 = x + y + z = x\left(1 + \frac{1}{2} + \frac{1}{3}\right) = \frac{11\,x}{6}$

$$x = \tfrac{6}{11}, \;\; y = \tfrac{3}{11}, \;\; z = \tfrac{2}{11}.$$

Case 2: $x = 0$, $y > 0$, and $z > 0$. $\lambda_2 = \lambda_3 = 0$

$$\lambda_0 = 4y = 6z, \quad \text{so } z = \tfrac{2y}{3}.$$

$$1 = y\left(1 + \tfrac{2}{3}\right) = \tfrac{5\,y}{3}, \quad y = \tfrac{3}{5}, \quad z = \tfrac{2}{3} \cdot \tfrac{3}{5} = \tfrac{2}{5}$$

$$\lambda_0 = 4y + \lambda_2 = 4\left(\tfrac{3}{5}\right) = \tfrac{12}{5}, \;\; \lambda_1 = \lambda_0 - 2x = \tfrac{12}{5} > 0.$$

$\left(0, \frac{3}{5}, \frac{2}{5}\right)$ is an allowable point for maximum.

Example 3, continued

Case 3: $y = 0,\ x > 0,\ z > 0. \quad \lambda_1 = \lambda_3 = 0$

$$\lambda_0 = 2x = 6z, \quad x = 3z.$$

$$1 = x + y + z = z(3+1), \quad z = \tfrac{1}{4},\ x = \tfrac{3}{4},\ \lambda_0 = \tfrac{3}{2}.$$

$$\lambda_2 = \lambda_0 - 4y = \tfrac{3}{2} > 0$$

$\left(\frac{3}{4}, 0, \frac{1}{4}\right)$ is an allowable point for a maximum.

Case 4: $z = 0, x > 0, y > 0 \quad \lambda_1 = \lambda_2 = 0$

$$\lambda_0 = 2x = 4y, \quad x = 2y.$$

$$1 = x + y + z = y(2+1), \quad y = \tfrac{1}{3} \quad x = \tfrac{2}{3} \quad \lambda_0 = 4y = \tfrac{4}{3}$$

$$\lambda_3 = \lambda_0 - 6z = \tfrac{4}{3} > 0$$

$\left(\frac{2}{3}, \frac{1}{3}, 0\right)$ is an allowable point for maximum.

Vertices $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ are possibilities. (2+1 tight)

Example 3, continued

Values of $f(x, y, z) = x^2 + 2y^2 + 3z^2$ at these points are as follows:

$$f\left(\tfrac{6}{11}, \tfrac{3}{11}, \tfrac{2}{11}\right) = \frac{36 + 18 + 12}{121} = \frac{66}{121} \approx 0.5454,$$

$$f\left(0, \tfrac{3}{5}, \tfrac{2}{5}\right) = \frac{18 + 12}{25} = \frac{30}{25} = 1.2,$$

$$f\left(\tfrac{3}{4}, 0, \tfrac{1}{4}\right) = \frac{9 + 3}{16} = \frac{12}{16} = 0.75,$$

$$f\left(\tfrac{2}{3}, \tfrac{1}{3}, 0\right) = \frac{4 + 2}{9} = \frac{6}{9} \approx 0.667,$$

$$f(1, 0, 0) = 1, \qquad f(0, 1, 0) = 2, \qquad f(0, 0, 1) = 3.$$

Maximum value of 3 is attained at $(0, 0, 1)$.

Minimum value of 0.5454 is attained at $\left(\frac{6}{11}, \frac{3}{11}, \frac{2}{11}\right)$

End of Example

Deficiencies of Nec KKT Theorem

Not easy to use Thm 3.12 to determine maximizers.

1. Need to verify that max exists: either $\mathscr{F}$ compact or $f(\mathbf{x})$ smaller near infinity.
2. Need to find all "critical points":
   (i) all points where constraint qualification fails
   (ii) all points that satisfy KKT-1,2,3.

   Not easy to show constraint qualification always holds, or
   find all points on boundary where constraint qualification fails.

Overcome by means of convexity and concavity.

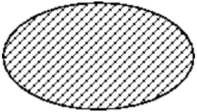
Convex constraints eliminates need for constraint qualification.

Concave (convex) objective fn insures that a KKT critical point is a global maximizer (minimizer).
Like an assumption on second derivative at all points of feasible set

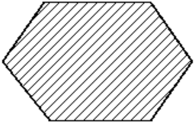
Definition

A set $\mathscr{D} \subset \mathbb{R}^n$ is called **convex** p.t.

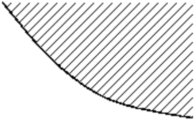
$$(1-t)\,\mathbf{x} + t\,\mathbf{y} \in \mathscr{D} \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathscr{D} \text{ and } 0 \leq t \leq 1.$$



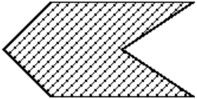
convex



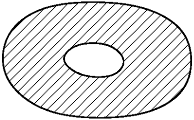
convex



convex



not convex



not convex

Concave/Convex Functions

Remark

Standard definitions of convex and concave functions

do not require C^1 and use only use values of function,

related to convex set.

Walker's definition of convex and concave functions assumes fn C^1

and gives a condition in terms of the tangent plane (p. 372).

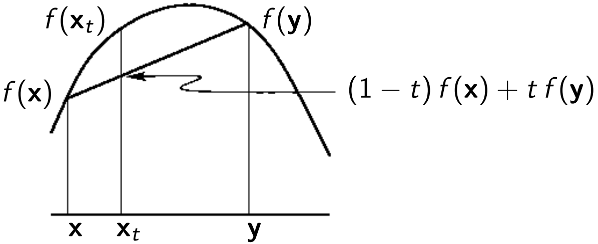
A theorem given later shows that our condition is equivalent

to Walker's for C^1 function.

Also see problem 7.5:5 in Walker

Use our defn in some proofs.

Concave Functions



Definition

$f : \mathscr{D} \subset \mathbb{R}^n \to \mathbb{R}$ is **concave on $\mathscr{D}$** p.t. for all $\mathbf{x}, \mathbf{y} \in \mathscr{D}$ & $0 \leq t \leq 1$,

$$\mathbf{x}_t = (1-t)\,\mathbf{x} + t\,\mathbf{y} \in \mathscr{D} \quad \& \quad f(\mathbf{x}_t) \geq (1-t)\,f(\mathbf{x}) + t\,f(\mathbf{y}).$$

equiv. to: **set of points below graph**,

$$\{(\mathbf{x}, y) \in \mathscr{D} \times \mathbb{R} : y \leq f(\mathbf{x})\}, \text{ is convex subset of } \mathbb{R}^{n+1}.$$

f is **strictly concave** p.t. for all $\mathbf{x}, \mathbf{y} \in \mathscr{D}$ with $\mathbf{x} \neq \mathbf{y}$ & $0 \leq t \leq 1$,

$$\mathbf{x}_t = (1-t)\,\mathbf{x} + t\,\mathbf{y} \in \mathscr{D} \quad \& \quad f(\mathbf{x}_t) > (1-t)\,f(\mathbf{x}) + t\,f(\mathbf{y}).$$

Convex Functions

Definition

$f : \mathscr{D} \subset \mathbb{R}^n \to \mathbb{R}$ is **convex** on $\mathscr{D}$, p.t. for all $\mathbf{x}, \mathbf{y} \in \mathscr{D}$ & $0 \le t \le 1$,

$$\mathbf{x}_t = (1-t)\,\mathbf{x} + t\,\mathbf{y} \in \mathscr{D} \quad \& \quad f(\mathbf{x}_t) \le (1-t)\,f(\mathbf{x}) + t\,f(\mathbf{y}).$$

equiv. to: **set of points above graph**,

$\{(\mathbf{x}, y) \in \mathscr{D} \times \mathbb{R} : y \ge f(\mathbf{x})\}$, is a convex set.

f is called **strictly convex** p.t. for all $\mathbf{x}, \mathbf{y} \in \mathscr{D}$ with $\mathbf{x} \ne \mathbf{y}$ & $0 < t < 1$,

$$\mathbf{x}_t = (1-t)\,\mathbf{x} + t\,\mathbf{y} \in \mathscr{D} \quad \& \quad f(\mathbf{x}_t) < (1-t)\,f(\mathbf{x}) + t\,f(\mathbf{y}).$$

Remark

Note that if f is either concave or convex on $\mathscr{D}$ then $\mathscr{D}$ is convex.

Continuity of Concave/Convex Functions

Theorem

If $f : \mathscr{D} \subset \mathbb{R}^n \to \mathbb{R}$ *is a concave or convex function on* $\mathscr{D}$*,*

then f *is continuous on* $\text{int}(\mathscr{D})$*.*

The proof is given in Sundaram "A First Course in Optimization Theory".

Since concave/convex functions are continuous,

reasonable to seek their maximum

Convexity of Feasible Set

Theorem

Assume that $\mathscr{D} \subset \mathbb{R}^n$ *is an open convex subset, and*

$$g_i : \mathscr{D} \to \mathbb{R} \text{ are } C^1 \text{ convex functions for } 1 \leq i \leq m.$$

Then, for any $\mathbf{b} \in \mathbb{R}^m$,

$$\mathscr{F}_{\mathbf{g},\mathbf{b}} = \{\, \mathbf{x} \in \mathscr{D} : g_i(\mathbf{x}) \leq b_i \;\; \text{for } 1 \leq i \leq m \,\} \text{ is a convex set.}$$

Proof.

Take $\mathbf{x}, \mathbf{y} \in \mathscr{F}_{\mathbf{g},\mathbf{b}}$ and let $\mathbf{x}_t = (1-t)\mathbf{x} + t\mathbf{y}$ for $0 \leq t \leq 1$.

For any $1 \leq i \leq m$,

$$g_i(\mathbf{x}_t) \leq (1-t)g_i(\mathbf{x}) + tg_i(\mathbf{y}) \leq (1-t)b_i + tb_i = b_i,$$

so $\mathbf{x}_t \in \mathscr{F}_{\mathbf{g},\mathbf{b}}$.

$\mathscr{F}_{\mathbf{g},\mathbf{b}}$ is convex. □

Slater Condition

Need a condition on feasible set $\mathscr{F}$ to insure that

an extremizer of f satisfies KKT-1,2.

Definition

Let $g_i : \mathscr{D} \to \mathbb{R}$ for $1 \leq i \leq m$,

$\mathscr{F}_{\mathbf{g},\mathbf{b}} = \{\mathbf{x} \in \mathscr{D} : g_i(\mathbf{x}) \leq b_i \text{ for } 1 \leq i \leq m\}$.

Constraint functions $\{ g_i \}$ satisfy **Slater Condition on $\mathscr{F}_{\mathbf{g},\mathbf{b}}$** p.t.

there exists $\bar{\mathbf{x}} \in \mathscr{F}_{\mathbf{g},\mathbf{b}}$ s.t. $g_i(\bar{\mathbf{x}}) < b_i$ for all $1 \leq i \leq m$.

If Slater Condition is satisfied then $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ has nonempty interior.

Very mild in comparison to constraint qualification.

Not needed to show soln of KKT-1,2 is an extremizer.

Theorem (**Karush-Kuhn-Tucker under Convexity**)

Assume $f, g_i : \mathbb{R}^n \to \mathbb{R}$ C^1 *for* $1 \le i \le m$,

$$\mathbf{x}^* \in \mathscr{F}_{\mathbf{g},\mathbf{b}} = \{\, \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \le b_i \ \text{ for } 1 \le i \le m \,\} \ \text{ for } \mathbf{b} \in \mathbb{R}^m.$$

a. *Assume* f *is concave.*

- **i.** *If* $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ *is convex and* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *satisfies KKT-1,2,3 with all* $\lambda_i^* \ge 0$, *then* f *has a maximum on* $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ *at* $\mathbf{x}^*$.
- **ii.** *If* f *has a maximum on* $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ *at* $\mathbf{x}^*$, *all* g_i *are convex, and* $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ *satisfies Slater condition, then there exist* $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_m^*) \ge \mathbf{0}$ *such that* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *satisfies KKT-1,2,3.*

b. *If* f *is convex rather than concave, then conclusions of part (a) are true with maximum replaced by minimum and*

$\lambda_i^* \ge 0$ *replaced by* $\lambda_i^* \le 0$.

(Assumptions on $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ *and* g_i *stay same.)*

Karush-Kuhn-Tucker Theorem, continued

Kuhn-Tucker published in 1951, and popularized result.

Karush thesis in 1939 had earlier result.

Fritz John has a related result in 1948.

Verify that $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ is convex by conditions on constraint functions.

If all $g_i(\mathbf{x})$ are convex then $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ is convex.

Later, allow rescaled convex function – still insures $\mathscr{F}_{\mathbf{g},\mathbf{b}}$ is convex.

In examples, once find a solution of KKT-1,2,3 then done.

Don't need to verify separately that max exists,

don't need Slater condition, or constraint qualification.

Give some further results about convexity and examples before proof.

Affine Function

Proposition

For $\mathbf{a} \in \mathbb{R}^n$ *and* $b \in \mathbb{R}$, *affine function on* $\mathbb{R}^n$ *given by*

$$g(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} + b = a_1 x_1 + \cdots + a_n x_n + b$$

is both concave and convex.

Proof.

$\mathbf{p}_0, \mathbf{p}_1 \in \mathbb{R}^n, \quad \mathbf{p}_t = (1-t)\,\mathbf{p}_0 + t\,\mathbf{p}_1.$

$$\begin{aligned} g(\mathbf{p}_t) &= \mathbf{a} \cdot [(1-t)\,\mathbf{p}_0 + t\,\mathbf{p}_1] + b \\ &= (1-t)\ [\mathbf{a} \cdot \mathbf{p}_0 + b] + t\ [\mathbf{a} \cdot \mathbf{p}_1 + b] \\ &= (1-t)\,g(\mathbf{p}_0) + t\,g(\mathbf{p}_1). \end{aligned}$$

have equality so both concave and convex. □

Second Derivative Test for Convexity/Concavity

Theorem

Let $\mathscr{D} \subset \mathbb{R}^n$ be open and convex and $f : \mathscr{D} \to \mathbb{R}$ be C^2.

a. *f is convex (resp. concave) on $\mathscr{D}$ iff $D^2 f(\mathbf{x})$ is positive semidefinite (resp. negative semidefinite) for all $\mathbf{x} \in \mathscr{D}$.*

b. *If $D^2 f(\mathbf{x})$ is positive (resp. negative) definite for all $\mathbf{x} \in \mathscr{D}$, then f is strictly convex (resp. strictly concave) on $\mathscr{D}$.*

Idea: If $D^2 f(\mathbf{x})$ is positive (resp. negative) definite,

then locally graph of f lies above (resp. below) tangent plane.

Proof makes this global.

Online Course Materials have proof based on Sundaram

Example 4

$$\begin{aligned} \text{Minimize :} \quad & f(x,y) = x^4 + y^4 + 12\,x^2 + 6\,y^2 - xy - x + y, \\ \text{Subject to :} \quad & g_1(x,y) = -x - y \leq \text{-}6 \\ & g_2(x,y) = -2x + y \leq \text{-}3 \\ & \text{-}x \leq 0,\ \text{-}y \leq 0. \end{aligned}$$

Constraints are linear and so are convex. $\mathscr{F}$ is convex

Objective function:

$$D^2 f(x,y) = \begin{bmatrix} 12\,x^2 + 24 & \text{-}1 \\ \text{-}1 & 12\,y^2 + 12 \end{bmatrix}.$$

$$12\,x^2 + 24 \geq 24 > 0 \quad \& \quad \det(D^2 f(x,y)) \geq 24(12) - 1 > 0,$$

so $D^2 f(x,y)$ is positive definite at all $\mathbf{x} \in \mathscr{F}$ and f is convex.

Example 4, continued

KKT-1 with $\lambda_1, \lambda_2, \mu_1, \mu_2$ multipliers for g_1, g_2, $-x \leq 0$, $-y \leq 0$.

$$0 = 4x^3 + 24x - y - 1 + \lambda_1 + 2\lambda_2 + \mu_1$$
$$0 = 4y^3 + 12y - x + 1 + \lambda_1 - \lambda_2 + \mu_2$$

If $x = 0$, then $y \leq -3$ so $-y > 0$. $-x < 0$ in $\mathscr{D}$, never tight $\mu_1 = 0$

$0 = \mu_2(-y)$: If $y = 0$, then $x = -g_1(x, 0) \geq 6$,

so $g_2(x, 0) = -2x \leq -12 < -3$, $\lambda_2 = 0$.

If $x > 6$, then $\lambda_1 = 0$, $0 = 4x^3 + 24 - 1$. $x = 6$, contradiction

For $x = 6$ & $y = 0$, 2nd equation gives

$0 = -6 + 1 + \lambda_1 + \mu_2$, or

$5 = \lambda_1 + \mu_2$.

Both these multipliers cannot be ≤ 0, so not minimum.

Example 4, continued

$x, y > 0$, so $\mu_2 = 0$. If both g_1 & g_2 tight then

$$6 = g_1(x, y) = x + y$$
$$3 = g_2(x, y) = 2x - y$$

solving yields $x = y = 3$.

If solution of KKT-1 then

$$0 = 4(3^4) + 24(3) - 3 - 1 + \lambda_1 + 2\lambda_2 = \lambda_1 + 2\lambda_2 + 176$$
$$0 = 4(3^3) + 12(3) - 3 + 1 + \lambda_1 - \lambda_2 = \lambda_1 - \lambda_2 + 140.$$

solving yields $\lambda_1 = \text{-}152 < 0$, $\lambda_2 = \text{-}12 < 0$.

$(x^*, y^*) = (3, 3)$, $\lambda_1 = \text{-}152$, $\lambda_2 = \text{-}12$, $\mu_1 = \mu_2 = 0$

satisfy KKT-1,2,3′ and is minimizer.

End of Example

Example 5: Lifetime of Equipment

Two machines with lifetimes x and y and costs \$1,600 and \$5,400,

average cost per year $\qquad \dfrac{1600}{x} + \dfrac{5400}{y}.$

Machine A has operating costs of $\$50j$ in j^{th} year

for average operating cost per year for lifetime of machine is

$$\frac{50 + 2(50) + \cdots + x(50)}{x} = \frac{50}{x} \cdot \frac{x(x+1)}{2} = 25(x+1).$$

Machine B has operating costs of $\$200j$ in j^{th} year

for average operating cost per year of

$$100(y+1).$$

Want combined total use of at least 20 years use, $x + y \geq 20$.

Example 5, continued

$$\begin{aligned} \text{Minimize :} \quad & f(x,y) = 25(x+1) + 100(y+1) + \frac{1600}{x} + \frac{5400}{y} \\ \text{Subject to :} \quad & g_1(x,y) = 20 - x - y \leq 0 \\ & g_2(x,y) = \text{-}x \leq 0 \\ & g_3(x,y) = \text{-}y \leq 0. \end{aligned}$$

Constraints are linear and convex, so $\mathscr{F}$ is convex.

$$D^2 f(x,y) = \begin{bmatrix} \frac{3200}{x^3} & 0 \\ 0 & \frac{10800}{y^3} \end{bmatrix} \text{ is positive definite all pts in } \mathbb{R}^2_{++},$$

so f is convex on $\mathbb{R}^2_{++}$.

f (cost) gets arbitrarily large near $x = 0$ or $y = 0$,

so minimum occurs for $x > 0$ and $y > 0$,

can ignore those multipliers

Example 5, continued

KKT-1,2 become

$$0 = 25 - \frac{1600}{x^2} + \lambda$$
$$0 = 100 - \frac{5400}{y^2} + \lambda$$
$$0 = \lambda(20 - x - y)$$

Assume constraint is effective and $y = 20 - x$.

first two equations give

$$\text{-}\lambda = 25 - \frac{1600}{x^2} = 100 - \frac{5400}{y^2}$$
$$0 = 75x^2y^2 - 5400x^2 + 1600y^2$$

$$0 = 75x^4 - 3000x^3 + 26200x^2 - 64000x + 640000$$

Example 5, continued

$$0 = 75x^4 - 3000x^3 + 26200x^2 - 64000x + 640000$$

has positive roots of $x \approx 12.07$ and 28.37.

If $x \approx 28.37$ then $y = 20 - x < 0$ so not feasible.

If $x \approx 12.07$, then $y \approx 7.93$ and $\lambda \approx -25 + \frac{1600}{12.07^2} \approx -14.02 < 0$.

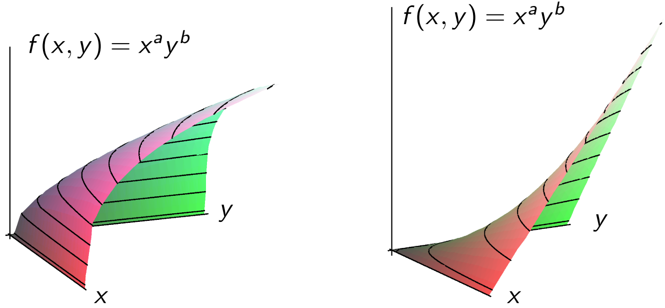
Minimizer.

Since a min exists on boundary, not necessary to check pts in interior;

However, if $\lambda = 0$, then $x = 8$ & $y = \sqrt{54} \approx 7.5$

$(8, \sqrt{54})$ is not in feasible set.

End of Example



$a = b = \frac{1}{3}$ $\qquad$ $a = b = 1$

Proposition

Let $f(x,y) = x^a y^b$ with $a, b > 0$.

a. *If $a + b \leq 1$, then f is concave and $-x^a y^b$ is convex on $\mathbb{R}^2_+$.*

b. *If $a + b > 1$, f is neither convex nor concave.*

Proof.

$$(x,y) \in \mathbb{R}^2_{++}, \quad D^2 f(x,y) = \begin{pmatrix} a(a-1)x^{a-2}y^b & abx^{a-1}y^{b-1} \\ abx^{a-1}y^{1-b} & b(b-1)x^a y^{b-2} \end{pmatrix}$$

$$a(a-1)x^{a-2}y^b < 0 \;\text{ if }\; a < 1$$

$$\det\left(D^2 f(x,y)\right) = ab(1-a-b)x^{2a-2}y^{2b-2}$$

$$\begin{cases} > 0 & \text{if } a+b<1 \\ = 0 & \text{if } a+b=1 \\ < 0 & \text{if } a+b>1. \end{cases}$$

If $a+b<1$, $D^2f(x,y)$ is neg. def., f is strictly concave on $\mathbb{R}^2_{++}$;

f is continuous on $\mathbb{R}^2_+$, so strictly concave on $\mathbb{R}^2_+ = \text{cl}(\mathbb{R}^2_{++})$.

If $a+b=1$, then $D^2f(x,y)$ is neg. semi-def. on $\mathbb{R}^2_{++}$

f is concave on $\mathbb{R}^2_+$;

If $a+b>1$, $D^2f(x,y)$ is indefinite and f neither concave nor convex. □

Cobb-Douglas Functions, contin.

$f(x, y) = x\,y$ is neither concave nor convex

(but it is rescaled concave function: discussed later)

Proposition

For $a_1 + \cdots + a_n < 1$ *and* $a_i > 0$ *for* $1 \leq i \leq n$.
$f(\mathbf{x}) = x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$ *concave on* $\mathbb{R}^n_+$ *and* $-x_1^{a_1} \cdots x_n^{a_n}$ *is convex.*
If $a_1 + \cdots + a_n > 1$, *then* f *is neither concave nor convex.*

Proof: $\mathbf{x} \in \mathbb{R}^n_{++}$,

$$f_{x_i} = a_i x_1^{a_1} \cdots x_i^{a_i - 1} \cdots x_n^{a_n}$$

$$f_{x_i x_i} = a_i(a_i - 1) x_1^{a_1} \cdots x_i^{a_i - 2} \cdots x_n^{a_n} = a_i(a_i - 1) x_i^{-2} f$$

$$f_{x_i x_j} = a_i a_j x_1^{a_1} \cdots x_i^{a_i - 1} \cdots x_j^{a_j - 1} \cdots x_n^{a_n} = a_i a_j x_i^{-1} x_j^{-1} f.$$

$$\Delta_k = \det \begin{bmatrix} a_1(a_1 - 1)x_1^{-2} f & \cdots & a_1 a_k x_1^{-1} x_k^{-1} f \\ \vdots & \ddots & \vdots \\ a_k a_1 x_k^{-1} x_1^{-1} f & \cdots & a_k(a_k - 1) x_k^{-2} f \end{bmatrix}$$

$$= a_1 \cdots a_k x_1^{-2} \cdots x_k^{-2} f^k \det \begin{bmatrix} a_1 - 1 & \cdots & a_k \\ \vdots & \ddots & \vdots \\ a_1 & \cdots & a_k - 1 \end{bmatrix}$$

$$= a_1 \cdots a_k x_1^{-2} \cdots x_k^{-2} f^k \, \bar{\Delta}_k \qquad \text{defines } \bar{\Delta}_k$$

Proof of Cobb-Douglas $\mathbb{R}^n$, continued

By induction $\bar{\Delta}_k = (\text{-}1)^k - (\text{-}1)^k(a_1 + \cdots + a_k)$.

$$\bar{\Delta}_1 = \det[a_1\text{-}1] = (\text{-}1)^1 - (\text{-}1)^1 a_1.$$

$$\bar{\Delta}_k = \det \begin{bmatrix} a_1\text{-}1 & \cdots & a_k \\ \vdots & \ddots & \vdots \\ a_1 & \cdots & a_k\text{-}1 \end{bmatrix}$$

$$= \det \begin{bmatrix} a_1\text{-}1 & \cdots & a_{k-1} & a_k \\ \vdots & \ddots & \vdots & \vdots \\ a_1 & \cdots & a_{k-1}\text{-}1 & a_k \\ a_1 & \cdots & a_{k-1} & a_k \end{bmatrix} + \det \begin{bmatrix} a_1\text{-}1 & \cdots & a_{k-1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ a_1 & \cdots & a_{k-1}\text{-}1 & 0 \\ a_1 & \cdots & a_{k-1} & \text{-}1 \end{bmatrix}$$

2nd det $= \text{-}\bar{\Delta}_{k-1} = \text{-}(\text{-}1)^{k-1} + (\text{-}1)^{k-1}(a_1 + \cdots + a_{k-1})$

$$= (\text{-}1)^k - (\text{-}1)^k\,(a_1 + \cdots + a_{k-1}).$$

By column operations on 1st det

$$\det \begin{bmatrix} a_1\text{-}1 & \cdots & a_{k-1} & a_k \\ \vdots & \ddots & \vdots & \vdots \\ a_1 & \cdots & a_{k-1}\text{-}1 & a_k \\ a_1 & \cdots & a_{k-1} & a_k \end{bmatrix} = a_k \begin{bmatrix} \text{-}1 & \cdots & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \text{-}1 & 1 \\ 0 & \cdots & 0 & 1 \end{bmatrix} = (\text{-}1)^{k-1} a_k$$

$$\begin{aligned} \bar{\Delta}_k &= (\text{-}1)^{k-1} a_k + (\text{-}1)^k - (\text{-}1)^k (a_1 + \cdots + a_{k-1}) \\ &= (\text{-}1)^k - (\text{-}1)^k (a_1 + \cdots + a_k). \\ \Delta_k &= (\text{-}1)^k a_1 \cdots a_k x_1^{\text{-}2} \cdots x_k^{\text{-}2} f^k \, (1 - a_1 - \cdots - a_k) \end{aligned}$$

Δ_k alternate signs as required for $D^2 f$ to be negative definite on $\mathbb{R}^n_{++}$,

f is strictly concave on $\mathbb{R}^n_{++}$. Since f is continuous,

f is concave on $\text{cl}(\mathbb{R}^n_{++}) = \mathbb{R}^n_{+}$. **QED**

Rescaling a Function

For KKT Thm, need $\mathscr{F}_{\mathbf{g},\mathbf{b}} = \{\, \mathbf{x} \in \mathscr{F} : g_i(\mathbf{x}) \leq b_i \quad \text{for all} \quad i \,\}$ convex.

By rescaling allow more fns than just convex fns.

Definition

$g : \mathscr{D} \subset \mathbb{R}^n \to \mathbb{R}$ is a **rescaling** of $\hat{g} : \mathscr{D} \subset \mathbb{R}^n \to \mathbb{R}$ p.t.

$\exists$ increasing function $\phi : \mathbb{R} \to \mathbb{R}$ s.t. $g(\mathbf{x}) = \phi \circ \hat{g}(\mathbf{x})$.

Since ϕ has an inverse, $\hat{g}(\mathbf{x}) = \phi^{-1} \circ g(\mathbf{x})$.

ϕ is a C^1 **rescaling** p.t. ϕ is C^1 and $\phi'(y) > 0$ for all $y \in \mathbb{R}$.

If $g(\mathbf{x})$ is a rescaling of a convex function $\hat{g}(\mathbf{x})$, then

$\mathbf{g}(\mathbf{x})$ is called a **rescaled convex function**.

Similarly, a **rescaled concave function**,

If $\mathbf{g}(\mathbf{x})$ is a C^1 rescaling, then $\mathbf{g}$ is

C^1 **rescaled convex function**, and C^1 **rescaled concave function**.

Rescaling Constraints

Theorem

If $g : \mathscr{D} \subset \mathbb{R}^n \to \mathbb{R}$ *is a rescaled convex function, then* $\mathscr{F}_{g,b} = \{\, \mathbf{x} \in \mathscr{D} : g(\mathbf{x}) \leq b \,\}$ *is convex for any* $b \in \mathbb{R}$.

Proof.

$g(\mathbf{x}) = \phi \circ \hat{g}(\mathbf{x})$ with $\phi : \mathbb{R} \to \mathbb{R}$ increasing and $\hat{g}$ convex.

$$\mathscr{F}_{g,b} = \{\, \mathbf{x} \in \mathscr{D} : \hat{g}(\mathbf{x}) \leq \phi^{-1}(b) \,\} \text{ is convex.}$$

□

A function is called **quasi-convex** p.t. all the sets $\mathscr{F}_{g,b}$ are convex.

(See Sundaram).

Showed a rescaled convex function is quasi-convex

All Cobb-Douglas Fns are Rescaled Concave Fns

Proposition

If $a_1, \dots, a_n > 0$, *then* $g(\mathbf{x}) = x_1^{a_1} \cdots x_n^{a_n}$ *is*

a C^1 *rescaled* C^1 *concave function on* $\mathbb{R}^n_{++}$.

-$g(\mathbf{x})$ *is a* C^1 *rescaling of a* C^1 *convex function on* $\mathbb{R}^n_{++}$.

Proof.

Let $b_i = \dfrac{a_i}{(2a_1 + \cdots + 2a_n)}$, for $1 \leq i \leq n$.

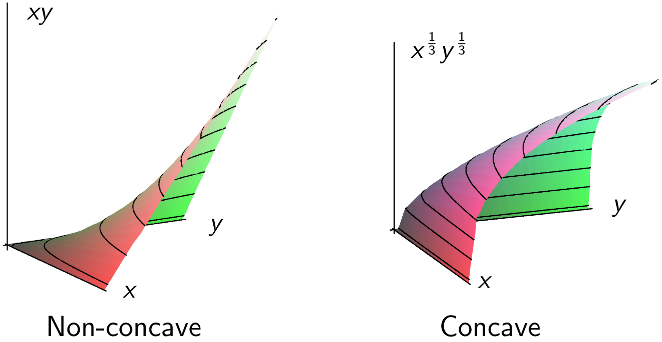
$b_1 + \cdots + b_n = \frac{1}{2} < 1$

$h(x, y, z) = x_1^{b_1} \cdots x_n^{b_n}$ is convex on $\mathbb{R}^n_+$.

$\phi(s) = s^{2a_1 + \cdots + 2a_n}$ is monotone on $\mathbb{R}_+$.

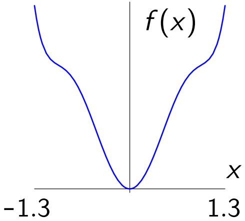
$g(\mathbf{x}) = \phi \circ h(\mathbf{x})$ is a C^1 rescaling on $\mathbb{R}^n_+$ by theorem. □

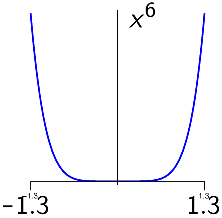
$f(x, y) = xy$ is a rescaled concave function, but not concave



Example of Rescaled Convex Function

$f(x) = x^6 - 2.9\,x^4 + 3\,x^2$





$f'(x) = x[6\,x^4 - 11.6\,x^2 + 6]$, single critical point $x = 0$.

$f''(x) = 30\,x^4 - 34.8\,x^2 + 6$, $f''(\pm 0.459) = 0 = f''(\pm 0.974)$.

$f''(x) < 0$ for $0.459 < x < 0.974$, $f(x)$ is not convex.

$f(x)$ is a rescaling of $\hat{f}(x) = x^6$ that is convex:

$\phi(y) = \left[f^{-1}(y)\right]^6$ satisfies $\phi \circ f(x) = \hat{f}(x)$.

$f(x)$ is a rescaled convex function that is not convex.

KKT with Rescaled Functions

Following corollary of KKT Theorem allows

rescaling of objective function as well as constraints.

Corollary

Assume that $f, g_i : \mathbb{R}^n \to \mathbb{R}$ *are* C^1 *for* $1 \leq i \leq m$,

each g_i *is a* C^1 *rescaled convex function*

f *is a* C^1 *rescaled concave (resp. convex) function*

$\mathbf{x}^* \in \mathscr{F}_{\mathbf{g},\mathbf{b}} = \{\, \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq b_i \;\text{ for } 1 \leq i \leq m \,\}$ *for* $\mathbf{b} \in \mathbb{R}^m$.

With these assumptions, the conclusions of the different parts of

KKT Theorem are valid.

Bazaraa et al allow f to be pseudo-concave.

Rescaled concave implies pseudo-concave.

See online class book.

Proof

$\hat{g}_i(\mathbf{x}) = \phi_i \circ g_i(\mathbf{x})$ with $\phi_i'(b_i) > 0$ for all $b_i \in \mathbb{R}$.

$\hat{f}(\mathbf{x}) = T \circ f(\mathbf{x})$ with $T'(y) > 0$ for all $y = f(\mathbf{x})$ with $\mathbf{x} \in \mathscr{F}$

Let $b_i' = \phi_i(b_i)$. If $g_i(\mathbf{x}^*) = b_i$ is tight, then $\hat{g}_i(\mathbf{x}^*) = b_i'$,

$$D\hat{g}_i(\mathbf{x}^*) = \phi_i'(b_i')Dg_i(\mathbf{x}^*) \quad \text{and} \quad D\hat{f}(\mathbf{x}^*) = T'(f(\mathbf{x}^*))\,Df(\mathbf{x}^*).$$

(a.i) $\mathscr{F}_{\mathbf{g},\mathbf{b}} = \{\, \mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq b_i \,\} = \{\, \mathbf{x} \in \mathbb{R}^n : \hat{g}_i(\mathbf{x}) \leq b_i' \,\}$ is convex.

If f satisfies KKT-1. then

$$D\hat{f}(\mathbf{x}^*) = T'(f(\mathbf{x}^*))\,Df(\mathbf{x}^*) = T'(f(\mathbf{x}^*))\sum_i \lambda_i\, Dg_i(\mathbf{x}^*),$$

so $\hat{f}$ satisfies KKT-1,2 with multipliers $T'(f(\mathbf{x}^*))\,\lambda_i > 0$.

By Theorem KKT(a.i), $\hat{f}$ has a maximum at $\mathbf{x}^*$.

Since T^{-1} is increasing, $f = T^{-1} \circ \hat{f}$ has a maximum at $\mathbf{x}^*$.

Proof, continued

(**a.ii**) If f has a maximum at $\mathbf{x}^*$, then since T is increasing,

$\hat{f}$ has a maximum at $\mathbf{x}^*$.

Applying Theorem KKT(a.ii) to $\hat{f}$ and $\hat{g}_i$ on $\mathscr{F}_{\mathbf{g},\mathbf{b}}$,

$$T'(f(\mathbf{x}^*))\, Df(\mathbf{x}^*) = D\hat{f}(\mathbf{x}^*) = \textstyle\sum_i \lambda_i\, D\hat{g}_i(\mathbf{x}^*) = \sum_i \lambda_i\, \phi_i'(b_i) Dg_i(\mathbf{x}^*),$$

using that $\lambda_i = 0$ unless $g_i(\mathbf{x}^*) = b_i$.

Since, $T'(f(\mathbf{x}^*)) > 0$ and $\phi_i'(b_i) > 0$ for all effective i,

conditions KKT-1.2 hold for f and the g_i

with multipliers $\dfrac{\lambda_i\, T_i'(b_i)}{T'(f(\mathbf{x}^*))} > 0$.

Example 6

$$\begin{aligned} \text{Maximize:} \quad & f(x,y,z) = xyz \\ \text{Subject to:} \quad & g_1(x,y,z) = 2x + y + 2z - 5 \le 0 \\ & g_2(x,y,z) = x + 2y + z - 4 \le 0, \\ & g_3(x,y,z) = \text{-}x \le 0, \\ & g_4(x,y,z) = \text{-}y \le 0, \qquad \text{and} \\ & g_5(x,y,z) = \text{-}z \le 0. \end{aligned}$$

All g_i are linear so convex

$\mathscr{F} = \{\, (x,y,z) \in \mathbb{R}^3 : g_i(x,y,z) \le 0 \text{ for } 1 \le i \le 5 \,\}$ is convex

f is C^1 rescaling of concave fn on $\mathbb{R}^3_+$

Could maximize $(xyz)^{\frac{1}{4}}$, but equations are more complicated.

$0 = f(0, y, z) = f(x, 0, z) = f(x, y, 0)$, so max in $\mathbb{R}^4_{++}$

g_i slack for $3 \leq i \leq 5$, so multipliers $\lambda_3 = \lambda_4 = \lambda_5 = 0$

KKT-1: $yz = 2\lambda_1 + \lambda_2$,

$xz = \lambda_1 + 2\lambda_2$,

$xy = 2\lambda_1 + \lambda_2$,

KKT-2: $0 = \lambda_1(5 - 2x - y - 2z)$,

$0 = \lambda_2(4 - x - 2y - z)$.

KKT-3: all $\lambda_i \geq 0$,

From 1st and 3rd equation, $yz = yx$, so $x = z$ (since $y > 0$)

Example 6, continued

If both g_1 and g_2 are effective,

$$5 = 4x + y$$
$$4 = 2x + 2y.$$

with solution $1 = x = y$, $z = x = 1$. $(1, 1, 1) \in \mathscr{F}$.

KKT-1 become

$$1 = 2\lambda_1 + \lambda_2$$
$$1 = \lambda_1 + 2\lambda_2$$

with solution $\lambda_1 = \lambda_2 = \frac{1}{3} > 0$.

f is rescaling of concave fn and all g_i are convex functions on $\mathbb{R}^3_{++}$.

$\mathbf{p}^* = (1, 1, 1)$ satisfies KKT-1,2 with $\lambda_1 = \lambda_2 = \frac{1}{3} > 0$.

By KKT Theorem, f must have a maximum on $\mathscr{F}$ at $\mathbf{p}^*$. **QED**

Slater Condition for Example 6

Don't need Slater condition to conclude that $\mathbf{p}^*$ is a maximizer.

Since $\bar{\mathbf{p}} = (0.5, 0.5, 0.5) \in \mathscr{F}$ has all $g_i(\bar{\mathbf{p}}) < 0$,

constraints satisfy Slater condition.

There are many other points with same property.

Slater Condition and Constraint Qualification

Remark

For previous example, constraint qualification is satisfied on $\mathscr{F}$

However, if add another constraint, $x + y + z - 3 \leq 0$,

$\mathbf{p}^*$ is still a solution of KKT-1,2,3

and a maximizer by KKT Theorem under convexity.

For this new example, there are three effective constraints at $\mathbf{p}^*$,

but the rank is still 2.

Does not satisfy the constraint qualification on $\mathscr{F}$.

Global Maximizers of Concave Functions

If $M = \max\{ f(\mathbf{x}) : \mathbf{x} \in \mathscr{F} \} < \infty$ exists,

$$\mathscr{F}^* = \{ \mathbf{x} \in \mathscr{F} : f(\mathbf{x}) = M \}$$

If max doesn't exist, $\mathscr{F}^* = \emptyset$.

Theorem

Assume that $f : \mathscr{F} \subset \mathbb{R}^n \to \mathbb{R}$ is concave. Then following hold:

- **a.** *Any local maximum point of f is a global maximum point of f.*
- **b.** *$\mathscr{F}^*$ is either empty or convex set.*
- **c.** *If f is strictly concave, then $\mathscr{F}^*$ is either empty or a single point.*

Proof.

(**a**) If not global max, then $\exists$ loc max $\mathbf{x}^*$ and $\mathbf{z} \neq \mathbf{x}^*$ s. t. $f(\mathbf{z}) > f(\mathbf{x}^*)$.

For $\mathbf{x}_t = (1-t)\,\mathbf{x}^* + t\,\mathbf{z}$ & $0 < t \leq 1$, $\mathbf{x}_t \in \mathscr{F}$

$$f(\mathbf{x}_t) \geq (1-t)f(\mathbf{x}^*) + t\,f(\mathbf{z}) > (1-t)f(\mathbf{x}^*) + t\,f(\mathbf{x}^*) = f(\mathbf{x}^*).$$

Since $f(\mathbf{x}_t) > f(\mathbf{x}^*)$ for small t, $\mathbf{x}^*$ cannot be a local max.

(**b**) Assume $\mathbf{x}_0, \mathbf{x}_1 \in \mathscr{F}^*$.

Let $\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\mathbf{x}_1$, & $M = \max\{\, f(\mathbf{x}) : \mathbf{x} \in \mathscr{F} \,\}$. Then

$$M \geq f(\mathbf{x}_t) \geq (1-t)f(\mathbf{x}_0) + t\,f(\mathbf{x}_1) = (1-t)M + tM = M,$$

$f(\mathbf{x}_t) = M$ & $\mathbf{x}_t \in \mathscr{F}^*$ for $0 \leq t \leq 1$.

$\mathscr{F}^*$ is convex.

(**c**) If $\mathbf{x}_0, \mathbf{x}_1 \in \mathscr{F}^*$ & $\mathbf{x}_0 \neq \mathbf{x}_1$, strict convexity implies $f(\mathbf{x}_t) > M$.

Contradiction implies at most one point. □

First Order Derivative Conditions for Convexity/Concavity

For C^1 fns on convex domain, our defn is equiv to Walker's

Theorem

Assume $\mathscr{D} \subset \mathbb{R}^n$ be open and convex and $f : \mathscr{D} \to \mathbb{R}$ is C^1.

a. *f is concave iff* $f(\mathbf{y}) \leq f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathscr{D}$.

b. *f is convex iff* $f(\mathbf{y}) \geq f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathscr{D}$.

Concavity iff every $\mathbf{p}$ graph of $f(\mathbf{x})$ lies below tangent plane at $\mathbf{p}$

$$f(\mathbf{x}) \leq f(\mathbf{p}) + Df(\mathbf{p})(\mathbf{x} - \mathbf{p})$$

Convexity iff every $\mathbf{p}$ graph of $f(\mathbf{x})$ lies above tangent plane at $\mathbf{p}$

$$f(\mathbf{x}) \geq f(\mathbf{p}) + Df(\mathbf{p})(\mathbf{x} - \mathbf{p})$$

Proof First Order Derivative Conditions

Proof: (a) ($\Rightarrow$) $\quad \mathbf{x}_t = (1-t)\mathbf{x} + t\mathbf{y} = \mathbf{x} + t(\mathbf{y}-\mathbf{x})$.

$$Df(\mathbf{x})(\mathbf{y}-\mathbf{x}) = \lim_{t\to 0+} \frac{f(\mathbf{x}_t) - f(\mathbf{x})}{t} \qquad \text{Chain Rule}$$

$$\geq \lim_{t\to 0+} \frac{(1-t)\,f(\mathbf{x}) + t\,f(\mathbf{y}) - f(\mathbf{x})}{t} \qquad \text{concave}$$

$$= \lim_{t\to 0+} \frac{t\,[f(\mathbf{y}) - f(\mathbf{x})]}{t} = f(\mathbf{y}) - f(\mathbf{x})$$

Proof First Order Derivative Conditions

$(\Leftarrow)$ Assume $Df(\mathbf{x})(\mathbf{y}-\mathbf{x}) \geq f(\mathbf{y}) - f(\mathbf{x})$

Let $\mathbf{x}_t = (1-t)\,\mathbf{x} + t\,\mathbf{y}$ and $\mathbf{w}_t = \mathbf{y} - \mathbf{x}_t = (1-t)(\mathbf{y}-\mathbf{x})$, so

$$\mathbf{x} - \mathbf{x}_t = -\left(\frac{t}{1-t}\right)\mathbf{w}_t.$$

$$f(\mathbf{x}) - f(\mathbf{x}_t) \leq Df(\mathbf{x}_t)(\mathbf{x}-\mathbf{x}_t) = -\left(\frac{t}{1-t}\right) Df(\mathbf{x}_t)\mathbf{w}_t \quad \text{and}$$

$$f(\mathbf{y}) - f(\mathbf{x}_t) \leq Df(\mathbf{x}_t)(\mathbf{y}-\mathbf{x}_t) = Df(\mathbf{x}_t)\mathbf{w}_t.$$

Multiplying first inequality by $(1-t)$, the second by t, and adding

$$(1-t)\,f(\mathbf{x}) + t\,f(\mathbf{y}) - f(\mathbf{x}_t) \leq 0, \qquad \text{or}$$

$$(1-t)\,f(\mathbf{x}) + t\,f(\mathbf{y}) \leq f(\mathbf{x}_t). \qquad f \text{ is concave}$$

QED

Directional Derivative Condition

Theorem

Assume that $f : \mathscr{F} \subset \mathbb{R}^n \to \mathbb{R}$ *is concave. and* $\mathbf{x}^* \in \mathscr{F}$*.*

$\mathbf{x}^*$ *maximizes* f *on* $\mathscr{F}$ *iff*

$$Df(\mathbf{x}^*)\,\mathbf{v} \le 0 \text{ for all vectors } \mathbf{v} \text{ that point into } \mathscr{F} \text{ at } \mathbf{x}^*.$$

Remark

For $\mathbf{x}^* \in \text{int}(\mathscr{F})$, it follows that $\mathbf{x}^*$ is a maximizer iff it is a critical point.

So result generalizes critical point condition.

Proof

Concave implies directional derivatives exist, so don't need C^1.

($\Rightarrow$) Assume $\mathbf{x}^*$ is maximizer and $\mathbf{v}$ points into $\mathscr{F}$ at $\mathbf{x}^*$.

For small $t \geq 0$, $\mathbf{x}^* + t\mathbf{v} \in \mathscr{F}$ and $f(\mathbf{x}^* + t\mathbf{v}) \leq f(\mathbf{x}^*)$.

$$Df(\mathbf{x}^*)\,\mathbf{v} = \lim_{t\to 0+} \frac{f(\mathbf{x}^* + t\mathbf{v}) - f(\mathbf{x}^*)}{t} \leq 0.$$

Proof, continued

($\Leftarrow$) Assume $Df(\mathbf{x}^*)\,\mathbf{v} \leq 0$ for all vectors $\mathbf{v}$ that point into $\mathscr{F}$ at $\mathbf{x}^*$.

If $\mathbf{x}^*$ is not a maximizer, then there exists $\mathbf{z} \in \mathscr{D}$ s.t. $f(\mathbf{z}) > f(\mathbf{x}^*)$.

$\mathbf{v} = \mathbf{z} - \mathbf{x}^*$ points into $\mathscr{F}$ at $\mathbf{x}^*$.

For $0 \leq t \leq 1$, $\mathbf{x}_t = (1-t)\,\mathbf{x}^* + t\,\mathbf{z} = \mathbf{x}^* + t\,\mathbf{v} \in \mathscr{F}$ and

$$f(\mathbf{x}_t) \geq (1-t)\,f(\mathbf{x}^*) + t\,f(\mathbf{z}) = f(\mathbf{x}^*) + t\,[f(\mathbf{z}) - f(\mathbf{x}^*)] \qquad \text{so}$$

$$Df(\mathbf{x}^*)\,\mathbf{v} = \lim_{t\to 0+} \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{t} \geq \lim_{t\to 0+} \frac{t\,[f(\mathbf{z}) - f(\mathbf{x}^*)]}{t}$$

$$= f(\mathbf{z}) - f(\mathbf{x}^*) > 0.$$

contradiction

$\mathbf{x}^*$ must be a maximizer. **QED**

Proof of KKT Under Convexity Theorem 3.17.a.i

Assume KKT-1,2,3 conditions at $(\boldsymbol{\lambda}^*, \mathbf{x}^*)$ $\lambda_i^* \geq 0$.

$\mathscr{F}$ is convex. f restricted to $\mathscr{F}$ is concave.

$\mathbf{E} = \mathbf{E}(\mathbf{x}^*)$ be effective constraints

$\mathbf{v}$ a vector that points into $\mathscr{F}$ at $\mathbf{x}^*$.

If $i \notin \mathbf{E}$, then $\lambda_i^* = 0$, and $\lambda_i^* Dg_i(\mathbf{x}^*)\,\mathbf{v} = 0$.

If $i \in \mathbf{E}$, then $g_i(\mathbf{x}^*) = b_i$, $g_i(\mathbf{x}^* + t\mathbf{v}) \leq b_i = g_i(\mathbf{x}^*)$,

$$\frac{g_i(\mathbf{x}^* + t\mathbf{v}) - g_i(\mathbf{x}^*)}{t} \leq 0 \qquad \text{for } t > 0, \text{ and}$$

$$Dg_i(\mathbf{x}^*)\,\mathbf{v} = \lim_{t\to 0+} \frac{g_i(\mathbf{x}^* + t\mathbf{v}) - g_i(\mathbf{x}^*)}{t} \leq 0.$$

Since $\lambda_i^* \geq 0$, $\lambda_i^* Dg_i(\mathbf{x}^*)\mathbf{v} \leq 0$.

$Df(\mathbf{x}^*)\mathbf{v} = \sum_i \lambda_i^* Dg_i(\mathbf{x}^*)\mathbf{v} \leq 0.$ if $\mathbf{v}$ points into $\mathscr{F}$.

By previous theorem, $\mathbf{x}^*$ is a maximizer. **QED**

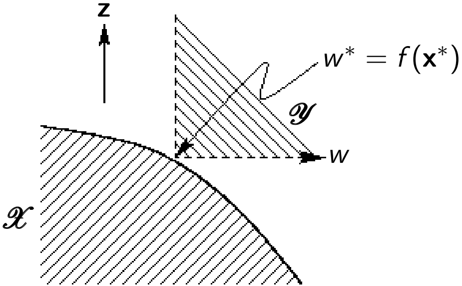
Proof of KKT Theorem a.ii

Assume that f has a maximum on $\mathscr{F}$ at $\mathbf{x}^*$.

For correct choice of $\lambda_i^* \geq 0$ show $\mathbf{x}^*$ is an **interior maximizer** of

$$L(\mathbf{x}, \boldsymbol{\lambda}^*) = f(\mathbf{x}) + \sum_i \lambda_i^* (b_i - g_i(\mathbf{x})) \qquad \text{with } \boldsymbol{\lambda}^* \text{ fixed}$$

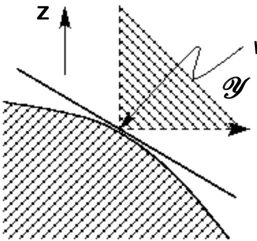
Use two disjoint convex sets in $\mathbb{R}^{m+1}$ space of values of constraints + 1



$$\mathscr{Y} = \{ (w, \mathbf{z}) \in \mathbb{R} \times \mathbb{R}^m : w > f(\mathbf{x}^*) \ \& \ \mathbf{z} \gg 0 \} \quad \text{is convex.}$$

$$\mathscr{X} = \{ (w, \mathbf{z}) \in \mathbb{R} \times \mathbb{R}^m : w \leq f(\mathbf{x}) \ \& \ \mathbf{z} \leq \mathbf{b} - g(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^n \}$$

Proof of Karush-Kuhn-Tucker, continued



$\mathscr{X}$ is shown to be convex. $\mathscr{X} \cap \mathscr{Y} = \emptyset$.

By convex separation theorem, $\exists\ (p, \mathbf{q}) \neq \mathbf{0}$ s.t.

$$p\,w + \mathbf{q} \cdot \mathbf{z} \leq p\,u + \mathbf{q} \cdot \mathbf{v} \qquad \text{for all } (w, \mathbf{z}) \in \mathscr{X},\ (u, \mathbf{v}) \in \mathscr{Y}. \quad (\dagger)$$

It is shown that $(p, \mathbf{q}) \geq 0$. Slater $\Rightarrow$ $p > 0$.

Any $\mathbf{x} \in \mathbb{R}^n$, $w = f(\mathbf{x})$ and $\mathbf{z} = \mathbf{b} - g(\mathbf{x})$,

$(u, \mathbf{v}) \in \mathscr{Y}$ converge to $(f(\mathbf{x}^*), \mathbf{0})$,

$$p\,f(\mathbf{x}) + \mathbf{q} \cdot (\mathbf{b} - g(\mathbf{x})) \leq p\,f(\mathbf{x}^*) \qquad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Proof of Karush-Kuhn-Tucker, continued

Setting $\boldsymbol{\lambda}^* = \left(\frac{1}{p}\right) \mathbf{q} = \left(\frac{q_1}{p}, \dots, \frac{q_m}{p}\right) \geq 0$, KKT-3.

$$f(\mathbf{x}) + \boldsymbol{\lambda}^* \cdot (\mathbf{b} - g(\mathbf{x})) \leq f(\mathbf{x}^*) \qquad \text{for all } \mathbf{x} \in \mathbf{U}.$$

$$L(\mathbf{x}, \boldsymbol{\lambda}^*) = f(\mathbf{x}) + \textstyle\sum_i \lambda_i^*(b_i - g_i(\mathbf{x})) \leq f(\mathbf{x}^*) \qquad \text{for all } \mathbf{x} \in \mathbf{U}.$$

For $\mathbf{x} = \mathbf{x}^*$, get $\sum_i \lambda_i^*(b_i - g_i(\mathbf{x}^*)) \leq 0$.

But $\lambda_i^* \geq 0$ and $b_i - g_i(\mathbf{x}^*) \geq 0$, so each

$$\lambda_i^*(b_i - g_i(\mathbf{x}^*)) = 0 \qquad \text{KKT-2}$$

$$L(\mathbf{x}, \boldsymbol{\lambda}^*) \leq f(\mathbf{x}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \qquad \text{for all } \mathbf{x} \in \mathbf{U}.$$

With $\boldsymbol{\lambda}^*$ fixed, $L(\mathbf{x}, \boldsymbol{\lambda}^*)$ has an interior maximum at $\mathbf{x}^*$ and

$$0 = D_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = Df(\mathbf{x}^*, \boldsymbol{\lambda}^*) - \textstyle\sum_i \lambda_i^* \, Dg_i(\mathbf{x}^*, \boldsymbol{\lambda}^*) \qquad \text{KKT-1.}$$

QED