# Optimization. A first course on mathematics for economists

Xavier Martinez-Giralt
Universitat Autònoma de Barcelona
xavier.martinez.giralt@uab.eu

II.3 Static optimization - Non-Linear programming

# Nonlinear programming - Definition

## Inequality restrictions

- $g_i(\mathbf{x}) \leq b_i, \; i = 1, \ldots, m$
- $g_i(\mathbf{x})$ continuous, continuously differentiable
- $b_i \in \mathbf{R}$

## Non-negativity restrictions

- $x_j \geq 0, \; j = 1, \ldots, n$

## Problem:

- $\max_{\mathbf{x}} f(\mathbf{x})$ s.t. $\begin{cases} \mathbf{g}(\mathbf{x}) \leq \mathbf{b} \\ \mathbf{x} \geq 0 \end{cases}$
- $f$ continuously differentiable.

# Nonlinear programming - Remarks

(i) No restriction on $m$ and $n$

(ii) The direction of the inequality in the restrictions $\leq$ is a convention.
e.g. $x_1 - 2x_2 \geq 7 \Leftrightarrow -x_1 + 2x_2 \leq -7$

(iii) An equality restriction can be rewritten as two inequality restrictions.
e.g. $x_1 - 2x_2 = 7 \Leftrightarrow x_1 - 2x_2 \leq 7$ and $-x_1 + 2x_2 \leq 7$

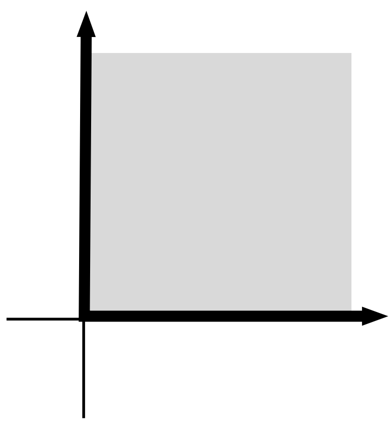
(iv) A free instrument $x_j$ can be rewritten as the difference of two non-negative instruments
e.g. $x_j = x_j' - x_j''$ with $x_j' \geq 0$ and $x_j'' \geq 0$.

(v) Consequence: Classical programming is a particular case of non-linear programming without non-negativity restrictions and the inequality restrictions written as equality restrictions.

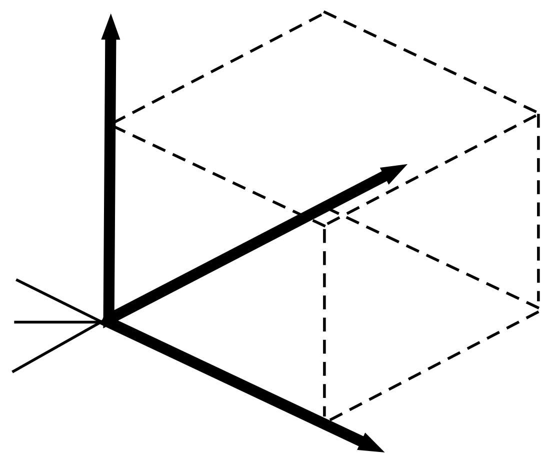
# Nonlinear programming - Geometry

- Each non-negativity restriction $x_j \geq 0$ defines a semi-space of non-negative values.
- The intersection of all non-negativity restrictions defines the non-negative orthant, a subset of the Euclidean $n$-dimensional space.


$x \in R,\ x \geq 0$



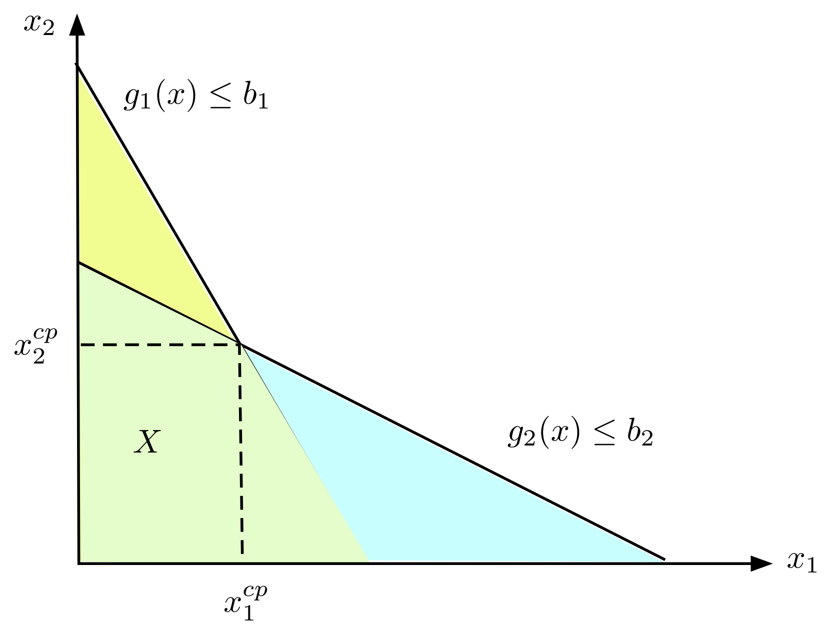
$x \in R^2,\ x \geq 0$



$x \in R^3,\ x \geq 0$

# Nonlinear programming - Geometry (2)

- Each inequality restriction $g_i(\mathbf{x}) \leq 0$ defines a set of points in $\mathbf{R}^n$.
- The intersection of all inequality restrictions defines a subset of the Euclidean $n$-dimensional space.
- The intersection of the non-negativity and inequality restrictions defines the feasible set, $X \subset \mathbf{R}^n$.



# Nonlinear programming - Geometry (3)

- Solution is a vector $\mathbf{x} \in X$ allowing to achieve the highest value of $f$.
- Given that $X$ is compact, assuming $f$ continuous allows to apply Weierstrass theorem so that a (set of) solution exists.
- Such solution(s) may be located either in the interior or on the frontier of $X$.
- Convexity assumptions are very important in non-linear programming problems
  - If $f$ is concave and all restrictions $g_i$ are convex, the local-global theorem tells us that a local maximum is also global and the set of solutions is a convex set.
  - Often this case is referred to as concave programming.

# A simple case

- Assume $m = 0$, so that only non-negativity restrictions.
- Problem reduces to $\max_{\mathbf{x}} f(\mathbf{x})$ s.t. $\mathbf{x} \geq 0$
- One way to solve the problem is using Taylor's expansion around a solution $\mathbf{x}^*$, assuming such a solution exists.
- Consider a neighborhood of $\mathbf{x}^* + \Delta\mathbf{x}$.
- Since $\mathbf{x}^*$ is a solution, $f(\mathbf{x}^*) \leq f(\mathbf{x}^* + h\Delta\mathbf{x})$, with $h \in \mathbf{R}$ arbitrarily small.
- Let $f$ be twice continuously differentiable. Then,
  $f(\mathbf{x}^* + h\Delta\mathbf{x}) =$
  $f(\mathbf{x}^*) + h\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}^*)\Delta\mathbf{x} + \frac{1}{2}h^2(\Delta\mathbf{x})'\frac{\partial^2 f}{\partial \mathbf{x^2}^2}(\mathbf{x}^* + \theta h\Delta\mathbf{x})\Delta\mathbf{x}$,
  with $\theta \in (0, 1)$.
- Substitution yields the fundamental inequality:
  $h\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}^*)\Delta\mathbf{x} + \frac{1}{2}h^2(\Delta\mathbf{x})'\frac{\partial^2 f}{\partial \mathbf{x^2}^2}(\mathbf{x}^* + \theta h\Delta\mathbf{x})\Delta\mathbf{x} \leq 0$

# A simple case (2)

- This is a necessary condition for a local maximum of $f$ at $\mathbf{x}^*$.
- If $\mathbf{x}^*$ is an interior solution $\mathbf{x}^* > 0$, the fundamental inequality has to be verified in every direction $\Delta\mathbf{x}$. This is equivalent to the same FOC of classical programming, $\frac{\partial f}{\partial x_j}(\mathbf{x}^*) = 0,\ \forall j$.
- Assume now $\exists j$ s.t. $x_j^* = 0$.
  - The fundamental inequality means that the only feasible direction is $\Delta x_j \geq 0$.
  - Then, dividing by $h$ and taking the $\lim_{h\to 0}$ we obtain, $\frac{\partial f}{\partial x_j}(\mathbf{x}^*)\Delta x_j \leq 0$.
  - Summarizing, we have that (i) if $x_j^* > 0$ then $\frac{\partial f}{\partial x_j}(\mathbf{x}^*) = 0$, and (ii) if $x_j^* = 0$ then $\frac{\partial f}{\partial x_j}(\mathbf{x}^*) \leq 0$.
  - Combining both conditions, it follows that $\frac{\partial f}{\partial x_j}(\mathbf{x}^*)x_j^* = 0$.

# A simple case (3)

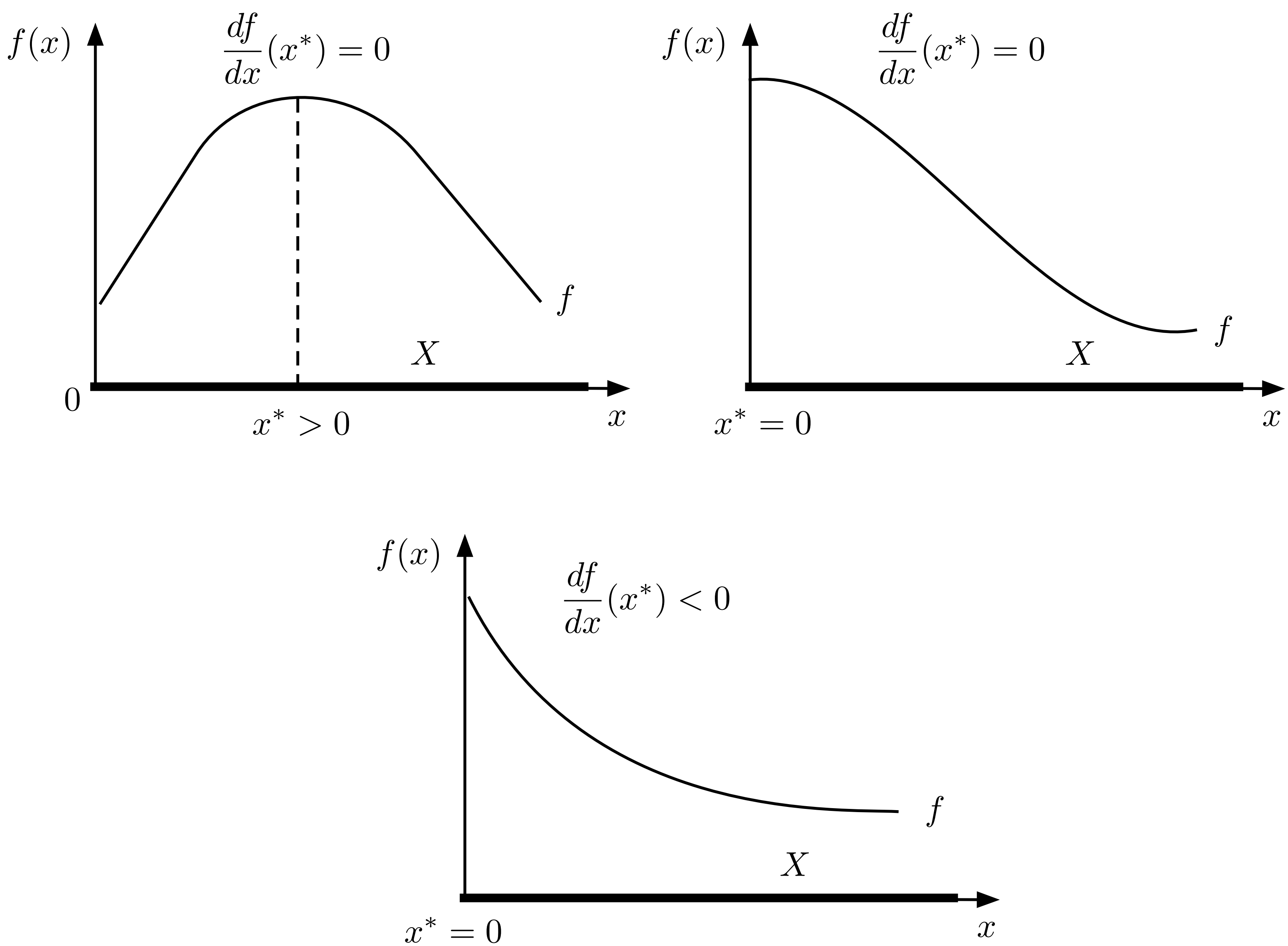
- Consider now the $n$-dimensions of the problem.
- $\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}^*)\mathbf{x}^* = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\mathbf{x}^*)x_j^* = 0$
- This condition says that the sum of the products cancels, but also that each element of the sum cancels given that $\mathbf{x} \geq 0$ and that the first partial derivatives are non-positive.
- Summarizing the FOCs are characterized by the following $2n+1$ conditions:

$$
\begin{gathered}
\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}^*) \leq 0 \\
\mathbf{x} \geq 0 \\
\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}^*)\mathbf{x}^* = 0
\end{gathered}
$$

- or equivalently $\forall j = 1, 2, \ldots, n$

$$
\begin{aligned}
\frac{\partial f}{\partial x_j}(\mathbf{x}^*) = 0, &\quad \text{if} \quad x_j^* > 0 \\
\frac{\partial f}{\partial x_j}(\mathbf{x}^*) \leq 0, &\quad \text{if} \quad x_j^* = 0
\end{aligned}
$$

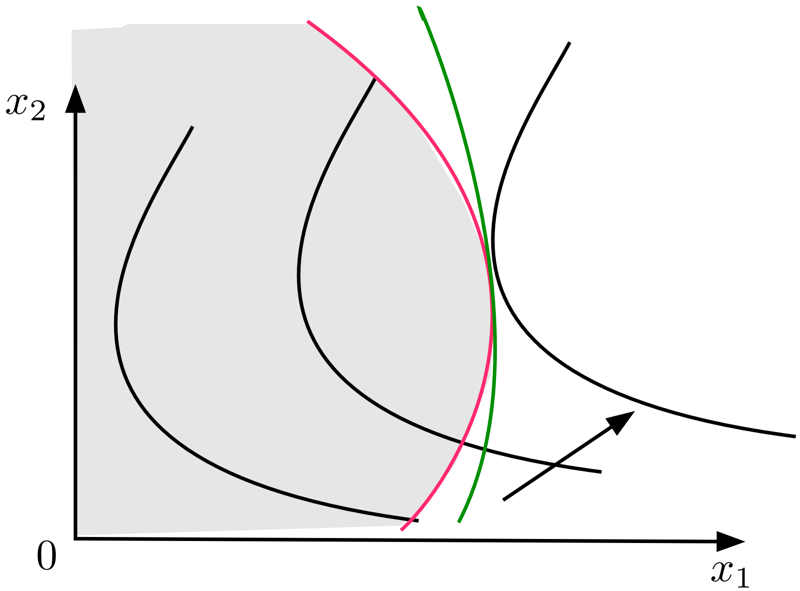
# Taxonomy of solutions



# The general case

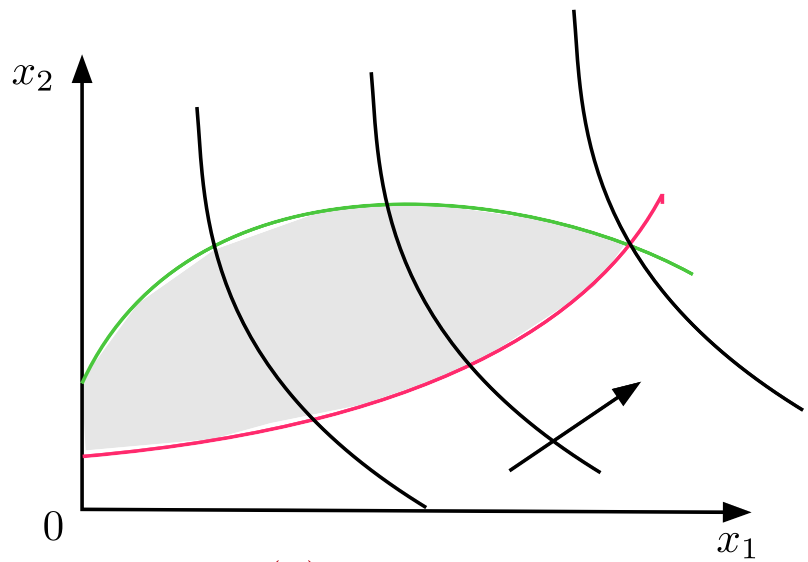
## Preliminaries

- The simple case provides the feeling for characterizing the solution of the general case.
- Constraints $g_i$ may be binding or not:




$g_1(x) \leq b_1$ binding

$g_2(x) \leq b_2$ not binding




$g_1(x) \leq b_1$ binding

$g_2(x) \leq b_2$ binding

# The general case (2)

## The problem

- $\max_{\mathbf{x}} f(\mathbf{x})$ s.t. $\begin{cases} \mathbf{g}(\mathbf{x}) \leq \mathbf{b} \\ \mathbf{x} \geq 0 \end{cases}$

## Strategy of solution - Solving the saddle-point problem

- Follow the logic of the classical programming problem
- Define a vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ of Lagrange multipliers, one for each inequality restriction $g_i$.
- Define the lagrangean function:
  $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda(\mathbf{b} - g(\mathbf{x}))$
- The set of FOC characterizing is known as the Kuhn-Tucker conditions

# Kuhn-Tucker conditions

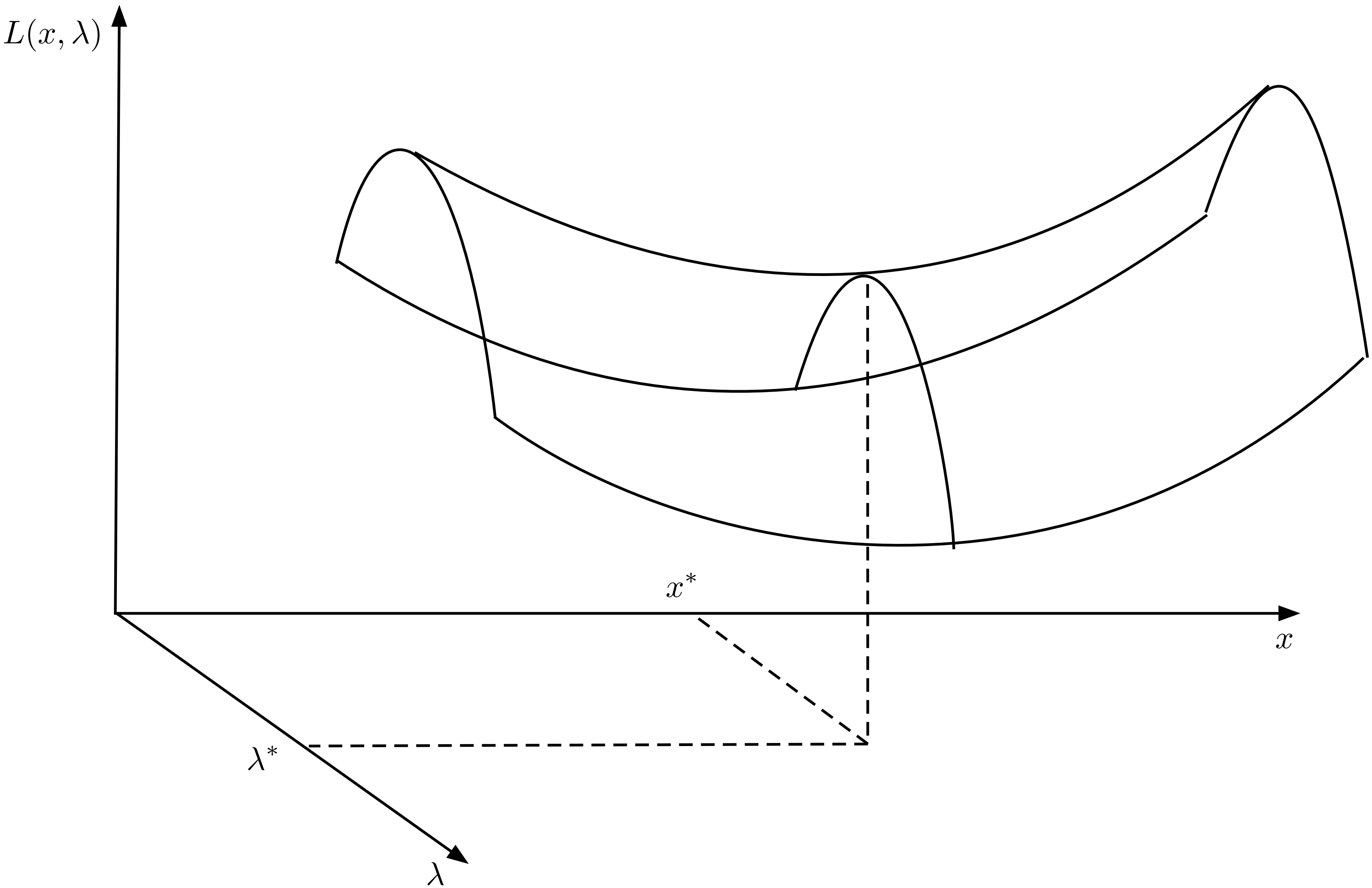
## The K-T conditions

$$\begin{aligned}
&\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}^*,\lambda^*) \le 0, && \frac{\partial L}{\partial \lambda}(\mathbf{x}^*,\lambda^*) \ge 0 \\
&\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}^*,\lambda^*)\mathbf{x}^* = 0, && \lambda^*\frac{\partial L}{\partial \lambda}(\mathbf{x}^*,\lambda^*) = 0 \\
&\mathbf{x}^* \ge 0, && \lambda^* \ge 0
\end{aligned}$$

## Remark

- Note the different sign of the partial derivatives wrt $\mathbf{x}$ and $\lambda \rightarrow$
- $\mathbf{x}^*$ maximizes $L$, while $\lambda^*$ minimizes $L$
- Thus $(\mathbf{x}^*,\lambda^*)$ is a saddle point of $L$, i.e.
  $L(\mathbf{x},\lambda^*) \le L(\mathbf{x}^*,\lambda^*) \le L(\mathbf{x}^*,\lambda),\ \forall \mathbf{x} \ge 0, \lambda \ge 0$
- Conditions content of Kuhn-Tucker theorem

# The saddle point problem



# Kuhn-Tucker conditions (2)

## Algebaric notation. The problem

$$\max_{x_1,\ldots,x_n} f(x_1,\ldots,x_n) \text{ subject to } \begin{cases} g_1(x_1,\ldots,x_n) \leq b_1 \\ \vdots \\ g_m(x_1,\ldots,x_n) \leq b_m \\ x_i \geq 0,\ i = 1,2,\ldots,n \end{cases}$$

## The K-T conditions

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} \leq 0, \qquad \frac{\partial L}{\partial \lambda_j} = b_j - g_j(\cdot) \geq 0$$

$$x_i \frac{\partial L}{\partial x_i} = x_i \left( \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} \right) = 0, \quad \lambda_j \frac{\partial L}{\partial \lambda_j} = \lambda_j (b_j - g_j(\cdot) = 0$$

$$x_i \geq 0,\ (i = 1,2,\ldots,n) \qquad \lambda_j \geq 0,\ (j = 1,2,\ldots,m)$$

# The Kuhn-Tucker theorem

(a) $\mathbf{x}^*$ solves the non-linear programming problem if $(\mathbf{x}^*, \lambda^*)$ is a solution of the saddle point problem.

(b) Under some conditions, $\mathbf{x}^*$ solves the non-linear programming problem only if $\exists \lambda^*$ for which $(\mathbf{x}^*, \lambda^*)$ solves the saddle point problem.

(c) What conditions?

- $f(\mathbf{x})$ is concave,
- $\forall j,\ g_j(\mathbf{x})$ are convex.
- *constraint qualification condition*: $\exists \mathbf{x}^\mathbf{0}$ such that $\mathbf{x}^\mathbf{0} \geq 0$ and $\mathbf{g}(\mathbf{x}^\mathbf{0}) < \mathbf{b}$.

# The Kuhn-Tucker theorem - Proof

## Proof of (a) - sufficiency ("if")

- If $(\mathbf{x}^*, \lambda^*)$ is a solution of the saddle point problem means
  $f(\mathbf{x}) + \lambda^*(\mathbf{b} - \mathbf{g}(\mathbf{x})) \leq f(\mathbf{x}^*) + \lambda^*(\mathbf{b} - \mathbf{g}(\mathbf{x}^*))$
  and
  $f(\mathbf{x}^*) + \lambda^*(\mathbf{b} - \mathbf{g}(\mathbf{x}^*)) \leq f(\mathbf{x}^*) + \lambda(\mathbf{b} - \mathbf{g}(\mathbf{x}^*))$
- Write the second inequality as
  $(\lambda - \lambda^*)(\mathbf{b} - \mathbf{g}(\mathbf{x}^*)) \geq 0, \ \forall \lambda \geq 0$
- Then,
  - for $\lambda$ such that $(\lambda - \lambda^*) > 0$, it follows that $\mathbf{b} - \mathbf{g}(\mathbf{x}^*) \geq 0$
  - for $\lambda$ such that $(\lambda - \lambda^*) < 0$, since $\mathbf{b} - \mathbf{g}(\mathbf{x}^*) \geq 0$ it follows that $\mathbf{b} - \mathbf{g}(\mathbf{x}^*) = 0$ and
  - in particular for $\lambda = 0$ it follows that $\lambda^*(\mathbf{b} - \mathbf{g}(\mathbf{x}^*)) = 0$
- Therefore, substituting in the first inequality,
  $f(\mathbf{x}) + \lambda^*(\mathbf{b} - \mathbf{g}(\mathbf{x})) \leq f(\mathbf{x}^*)$ implying $f(\mathbf{x}) \leq f(\mathbf{x}^*)$.

# The Kuhn-Tucker theorem - Proof (2)

## Proof of (b) - necessity ("only if")

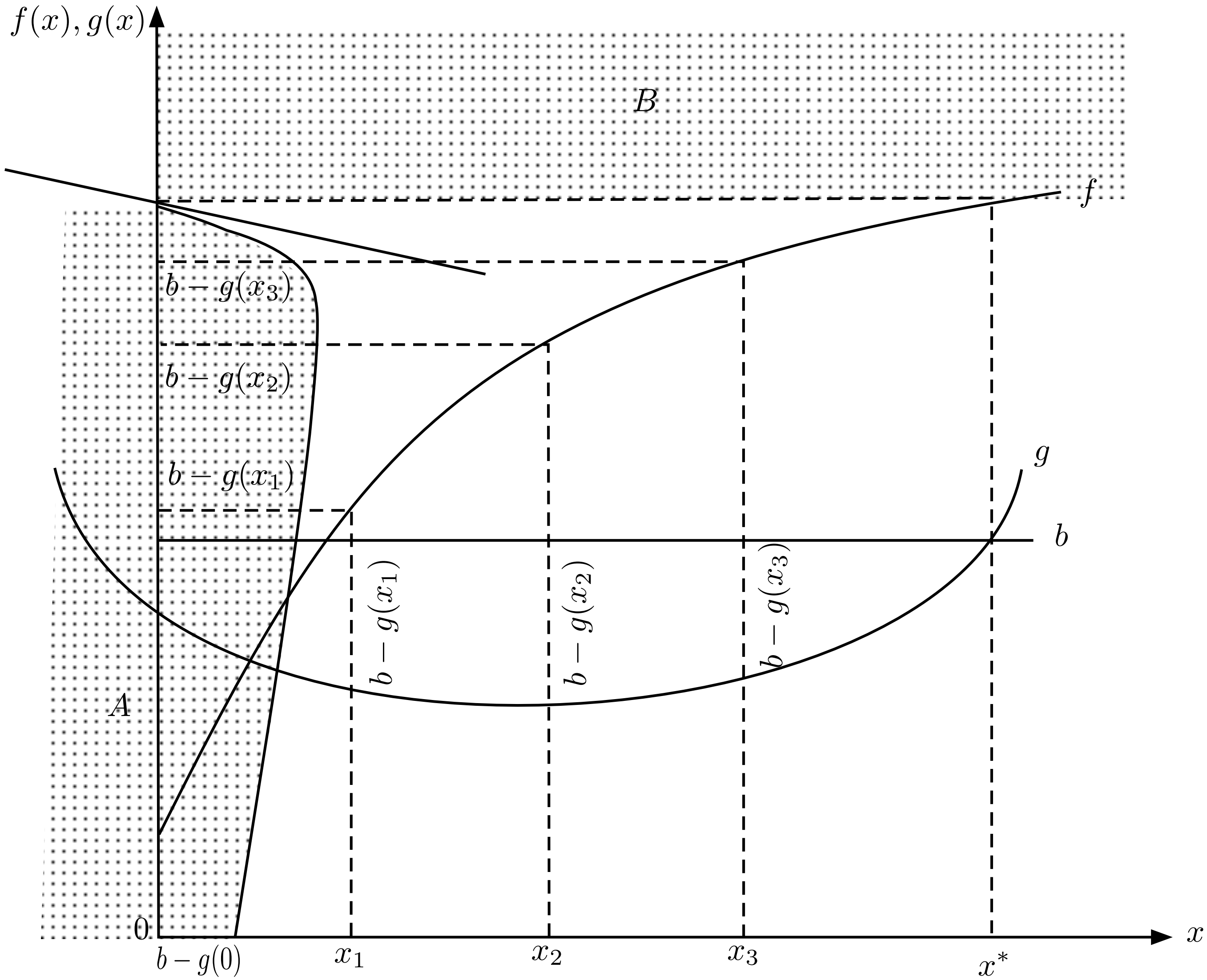
- Assume $\mathbf{x}^*$ solves the non-linear programming problem, i.e. $\mathbf{x}^* \geq 0,\ \mathbf{g}(\mathbf{x}^*) \leq \mathbf{b},$ and $f(\mathbf{x}^*) \geq f(\mathbf{x}),\ \forall \mathbf{x} \geq 0,\ \mathbf{g}(\mathbf{x}) \leq \mathbf{b}.$
- Let $a_0 \in \mathbf{R},\ b_0 \in \mathbf{R}$. Let $\mathbf{a} \in \mathbf{R}^m, \mathbf{b} \in \mathbf{R}^m$.
- Define the following $(m+1)$-dimensional convex sets:

$$A = \left\{ \begin{pmatrix} a_0 \\ \mathbf{a} \end{pmatrix} \middle| \begin{pmatrix} a_0 \\ \mathbf{a} \end{pmatrix} \leq \begin{pmatrix} f(\mathbf{x}) \\ \mathbf{b} - \mathbf{g}(\mathbf{x}) \end{pmatrix} \right\}$$

$$B = \left\{ \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \middle| \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} > \begin{pmatrix} f(\mathbf{x}^*) \\ \mathbf{0} \end{pmatrix} \right\}$$

- Let $m = n = 1$ Then,
  - $A$ is a set bounded by points with vertical distance $f(x)$ and horizontal distance $b - g(x)$. Thus convex.
  - $B$ is the interior of the quadrant with vertex at the point $(f(x^*), 0)$. Also convex.

# The sets $A$ and $B$



# The Kuhn-Tucker theorem - Proof (3)

## Proof of (b) - necessity ("only if") - [cont'd]

- Note that $A$ and $B$ are disjoint
- Applying the theorem on the separating hyperplane, there is a vector $(\lambda_0, \lambda)$, with $\lambda_0 \in \mathbf{R}$ and $\lambda \in \mathbf{R}^m$ such that
$$(\lambda_0, \lambda) \begin{pmatrix} a_0 \\ \mathbf{a} \end{pmatrix} \leq (\lambda_0, \lambda) \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix}, \ \forall \begin{pmatrix} a_0 \\ \mathbf{a} \end{pmatrix} \in A, \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \in B.$$
- Note that from the definition of $B$, the vector $(\lambda_0, \lambda)$ is nonnegative.
- Also, since $(f(\mathbf{x}^*), \mathbf{0})'$ is on the boundary of $A$, it follows that $\lambda_0 f(\mathbf{x}) + \lambda(\mathbf{b} - \mathbf{g}(\mathbf{x})) \leq \lambda_0 f(\mathbf{x}^*), \ \forall \mathbf{x} \geq \mathbf{0}$
- Because of the constraint qualification condition $\lambda_0 > 0$.
  - If $\lambda_0 = 0 \rightarrow \lambda(\mathbf{b} - \mathbf{g}(\mathbf{x})) \leq 0, \ \forall \mathbf{x} \geq \mathbf{0}$ and the nonnegativity of $\lambda$ would contradict the existence of an $\mathbf{x}^\mathbf{0} \geq \mathbf{0}$ such that $\mathbf{g}(\mathbf{x}^\mathbf{0}) < \mathbf{b}$.

# Separating hyperplane - recall

- $(\lambda_0, \lambda)$ separating hyperplane means:

  $$(\lambda_0, \lambda) \begin{pmatrix} a_0 \\ \mathbf{a} \end{pmatrix} \leq \begin{pmatrix} k_0 \\ \mathbf{k} \end{pmatrix} \text{ and}$$

  $$(\lambda_0, \lambda) \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix} \geq \begin{pmatrix} k_0 \\ \mathbf{k} \end{pmatrix}$$

- Hence,

  $$(\lambda_0, \lambda) \begin{pmatrix} a_0 \\ \mathbf{a} \end{pmatrix} \leq (\lambda_0, \lambda) \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix}$$

# The Kuhn-Tucker theorem - Proof (4)

## Proof of (b) - necessity ("only if") - [cont'd (2)]

- If $\lambda_0 > 0$ dividing both sides, we obtain $f(\mathbf{x}) + \lambda^*(\mathbf{b} - \mathbf{g}(\mathbf{x})) \leq f(\mathbf{x}^*)$, $\forall \mathbf{x} \geq \mathbf{0}$ with $\lambda^* = 1/\lambda_0$. $[KT1]$
- In particular, if $\mathbf{x} = \mathbf{x}^*$ it follows that $\lambda^*(\mathbf{b} - \mathbf{g}(\mathbf{x}^*)) \leq 0$
- But we know that $(\mathbf{b} \geq \mathbf{g}(\mathbf{x}^*))$ and $\lambda^* \geq \mathbf{0}$. Thus, $\lambda^*(\mathbf{b} - \mathbf{g}(\mathbf{x}^*)) = \mathbf{0}$. $[KT2]$
- Finally, define the Lagrangian as: $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda(\mathbf{b} - \mathbf{g}(\mathbf{x})$
- From $[KT1]$ and $[KT2]$ and from the assumption $\mathbf{y} \geq 0$, it follows that $(\mathbf{x}^*, \lambda^*)$ is a saddle point for $L(\mathbf{x}, \lambda)$ for $\mathbf{x} \geq \mathbf{0}, \lambda \geq \mathbf{0}$ thus proving the necessity part of the theorem.

## Summarizing

- Under the above assumptions $\mathbf{x}^*$ solves the nonlinear programming problem if ans only if $\exists \lambda^*$ such that $(\mathbf{x}^*, \lambda^*)$ solves the saddle point problem.

# The saddle-point problem - Remark

Additional assumption: $f(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are differentiable functions.
Part 1: $\max_{\mathbf{x}} L(\mathbf{x}, \lambda^*)$ - Conditions

$$\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}^*, \lambda^*) =\leq 0,$$
$$\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}^*, \lambda^*)\mathbf{x}^* = 0,$$
$$\mathbf{x}^* \geq 0.$$

Part 2: $\min_{\lambda} L(\mathbf{x}^*, \lambda)$ - Conditions

$$\frac{\partial L}{\partial \lambda}(\mathbf{x}^*, \lambda^*) \geq 0,$$
$$\lambda^* \frac{\partial L}{\partial \lambda}(\mathbf{x}^*, \lambda^*) = 0,$$
$$\lambda^* \geq 0.$$

# Geometry of the Kuhn-Tucker conditions

## The problem

$$\max_{x_1,x_2} f(x_1,x_2) \text{ subject to } \begin{cases} g_1(x_1,x_2) \leq b_1 \\ g_2(x_1,x_2) \leq b_2 \\ x_1, x_2 \geq 0 \end{cases}$$

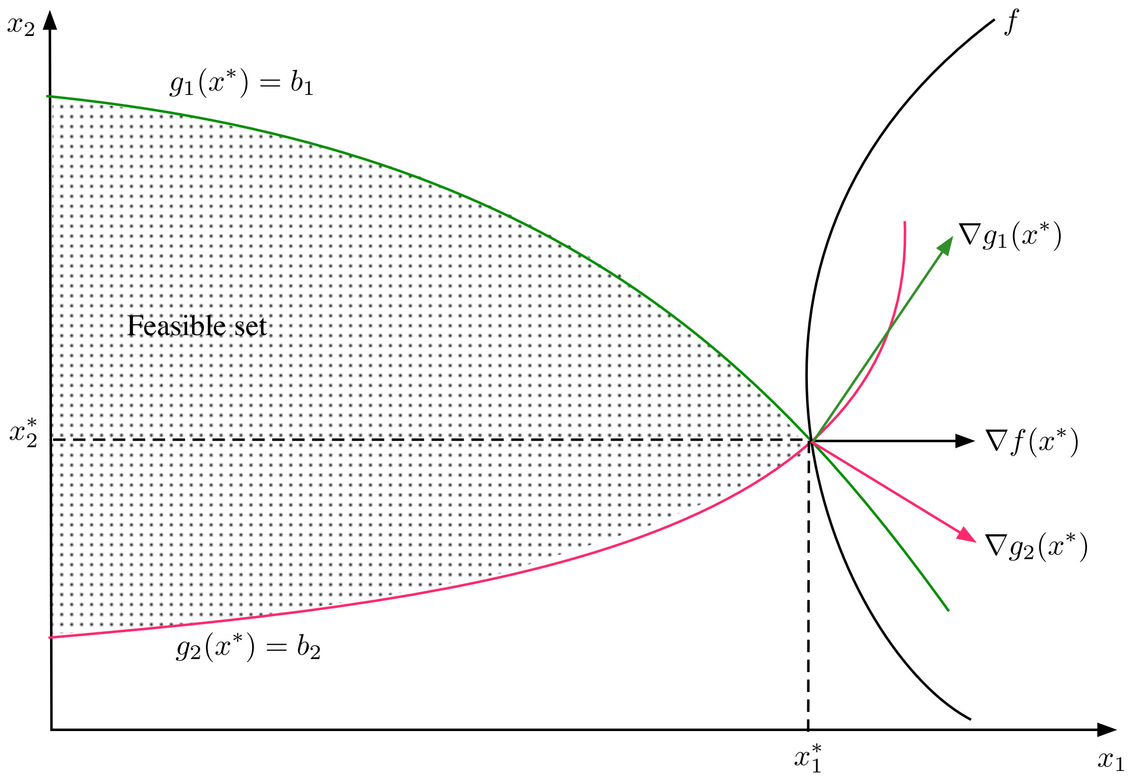
## The K-T conditions

- Let $x^* = (x_1^*, x_2^*)$ be an interior solution to this problem.
- K-T theorem says that $\nabla f(x*)$ must lie in the cone formed by the gradients of the restrictions $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$ ...
- ... i.e. $\nabla f(x^*)$ is a non-negative linear combination of $\nabla g_1(x^*)$ and $\nabla g_2(x^*)$
- Formally, $\exists(\lambda_1,\lambda_2) \geq 0$ s.t. $\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2$.

# Geometry of the Kuhn-Tucker conditions (2)

- Case 1
  - Case 1a: all restrictions as equalities and active, interior solution.
  - Case 1b: all restriction as equalities, some active, interior solution
- Case 2
  - Case 2a: some restrictions inactive, interior solution
  - Case 2b: some restrictions inactive, corner solution

# Case 1a: K-T conditions satisfied at $x^* > 0, g(x^*) = b$

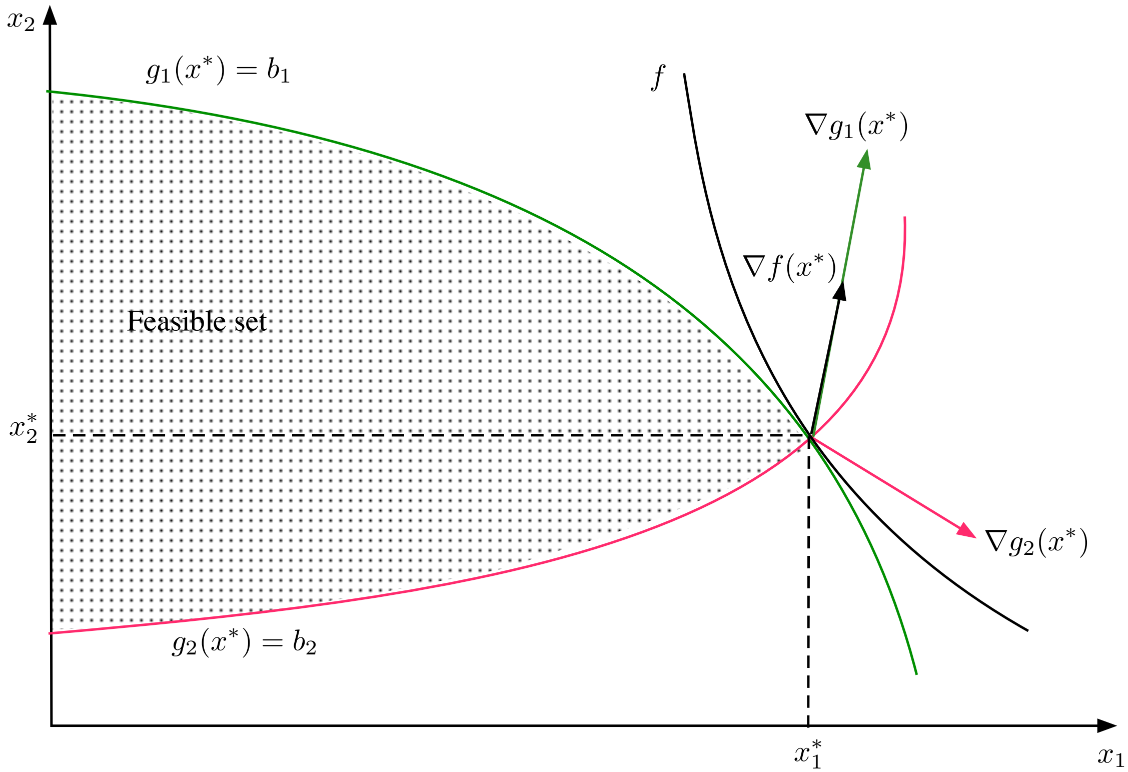



Remarks:

$g_1(x^*) = b_1, \; g_2(x^*) = b_2$

$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2$, with $\lambda_1 > 0$ and $\lambda_2 > 0$

# Case 1b: K-T conditions satisfied at $x^* > 0, g(x^*) = b$




Remarks:

$g_1(x^*) = b_1, g_2(x^*) = b_2$

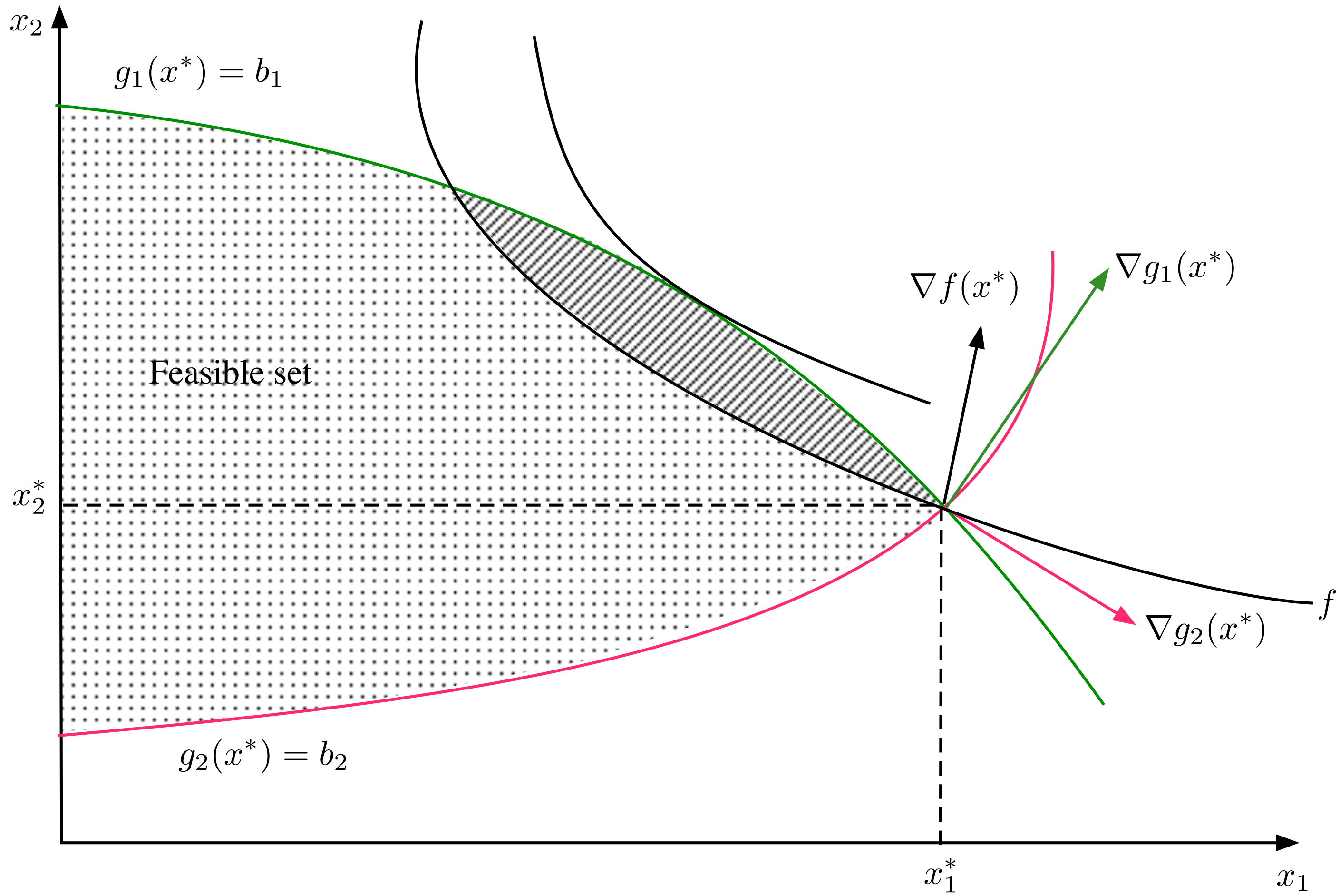
$\nabla f = \lambda_1 \nabla g_1$, with $\lambda_1 > 0$ and $\lambda_2 = 0$

# Geometry of the Kuhn-Tucker conditions (2)

## What if $\nabla f(x*)$ is not in the cone?

- Consider $f$ slightly perturbed (as in next figure) so that $\nabla f$ is no longer in the cone.
- (we could have perturbed $g_1$ and/or $g_2$ instead)
- Observe that now there is a “lens” between the contours of $f$ and $g_1$.
- This lens contains feasible points allowing to achieve large values of $f$ than $f(x^*)$.
- Thus $x^*$ is no longer a maximizer of $f$, as it does not satisfy K-T conditions.

# Geometry of the Kuhn-Tucker conditions (3)



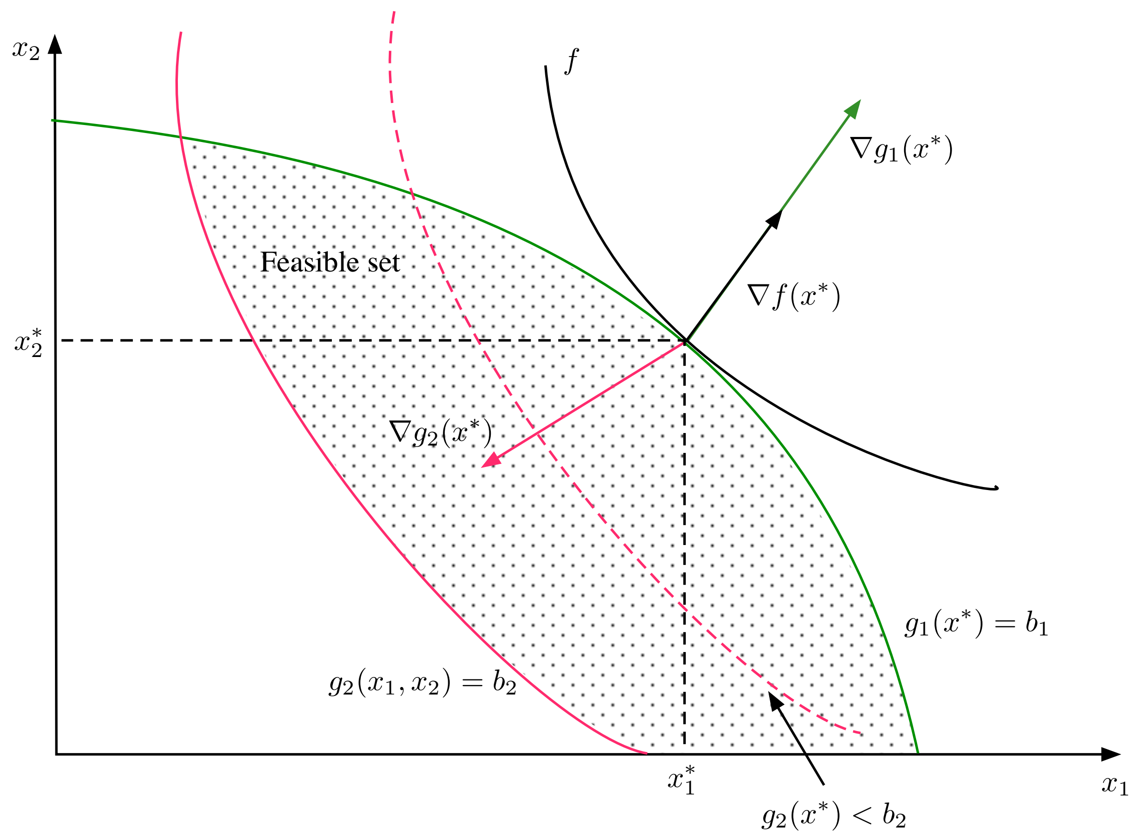
# Geometry of the Kuhn-Tucker conditions (4)

## Remark

- The *constraint qualification condition* precisely means that $\nabla f$ lies in the cone...
- ... i.e. that the restriccions are linearly independent

# Case 2a: K-T conditions satisfied at $x_j^* > 0, g_i(x^*) < b_i$

- Let $g_2(x^*) < b_2$. K-T conditions require $\lambda_2 = 0$, and $\nabla f$ lies in the cone defined by $g_1$.

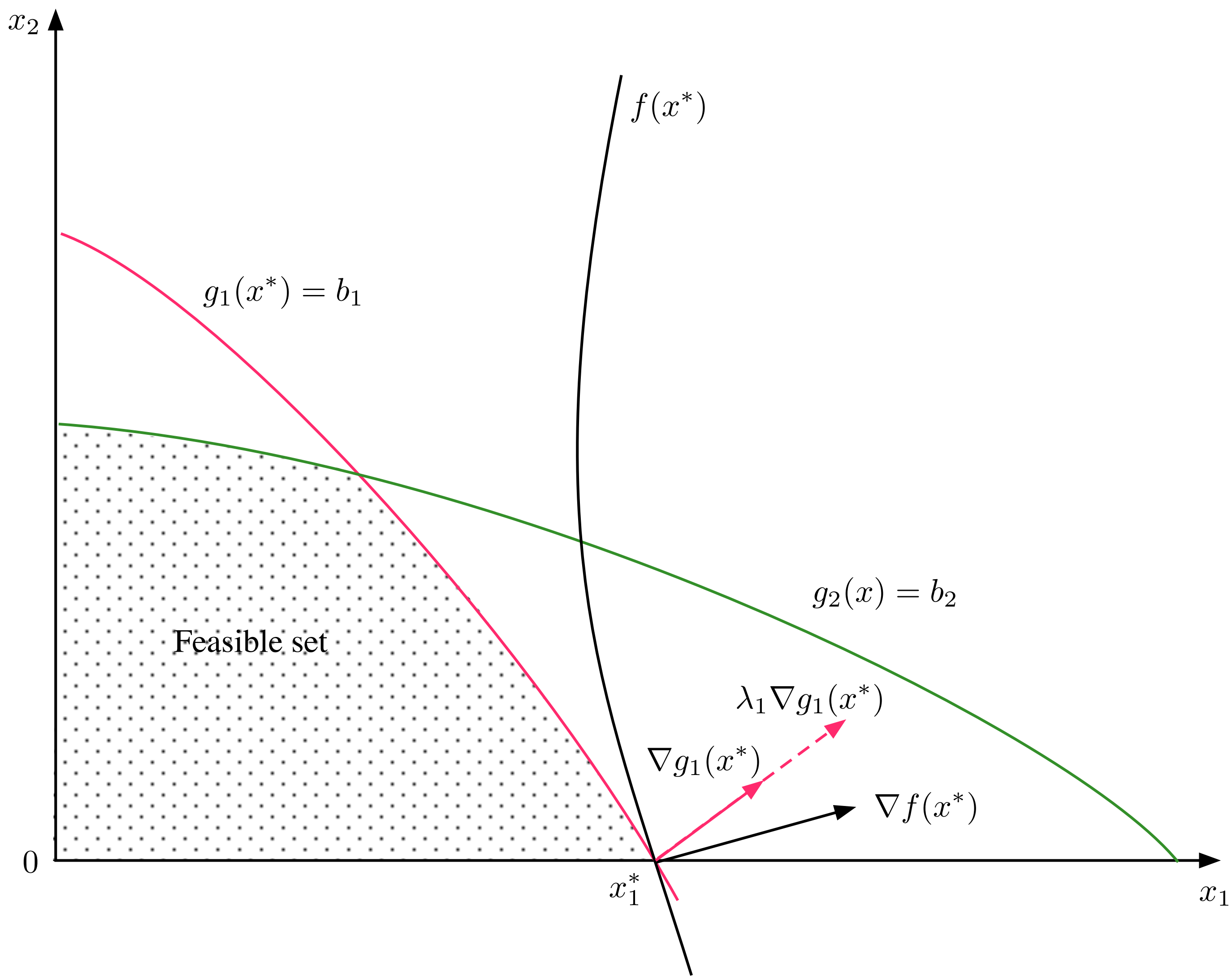



- $g_1(x^* = b_1, g_2(x^*) < b_2 \rightarrow \lambda_2 = 0; \nabla f = \lambda_1 \nabla g_1$

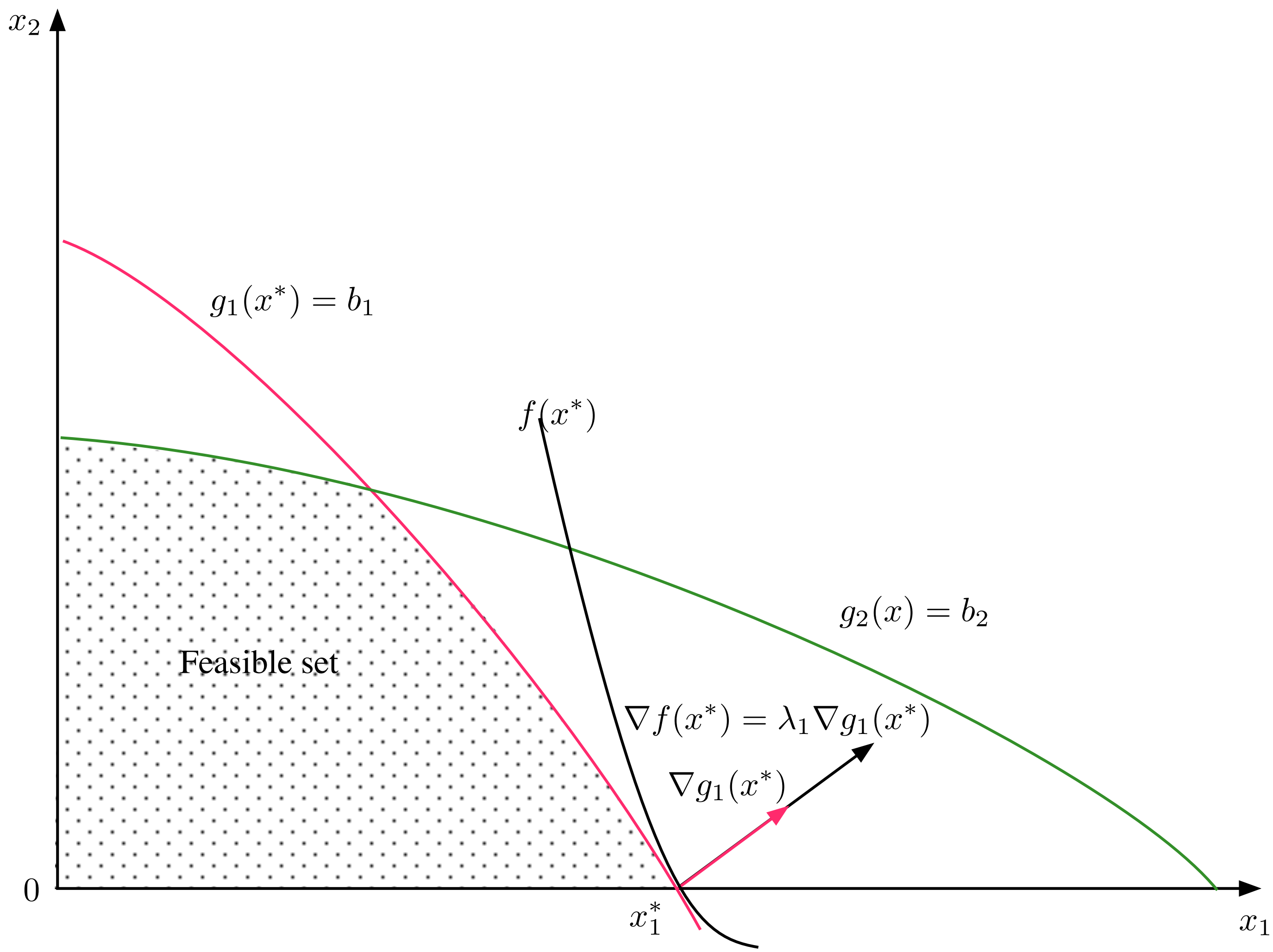
# Case 2b: K-T conditions satisfied at $x_j^* = 0, g_i(x^*) < b_i$

- Let $\frac{\partial f}{\partial x_2} < \sum_{i=1}^{2} \lambda_i \frac{\partial g_i}{\partial x_2}$. K-T conditions require $x_2^* = 0$.
- Assume $g_2(x^*) < b_2$ so that $\lambda_2 = 0$
- Accordingly,
  $\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1}(x^*) - \lambda_1 \frac{\partial g_1}{\partial x_1}(x^*) \leq 0$ and
  $\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2}(x^*) - \lambda_1 \frac{\partial g_1}{\partial x_2}(x^*) < 0$
- In the next figure it happens that $\frac{\partial f}{\partial x_1}(x^*) < \lambda_1 \frac{\partial g_1}{\partial x_1}(x^*)$, i.e. $\nabla f$ is not in the cone $\nabla g$.
- but it might be possible to have $\frac{\partial f}{\partial x_1}(x^*) = \lambda_1 \frac{\partial g_1}{\partial x_1}(x^*)$ in case of tangency at $x^*$.

# Case 2b: K-T conditions satisfied at $x_j^* = 0, g_i(x^*) < b_i$ (2)



# Case 2b: K-T conditions satisfied at $x_j^* = 0, g_i(x^*) < b_i$ (3)



# The envelope theorem

## Motivation

- Consider the problem $\max_x f(x)$ s.t. $g(x) \leq c$
- Generalize this problem by allowing $f$ and $g$ to depend on a parameter $\theta$, i.e. $\max_x f(x, \theta)$ s.t. $g(x, \theta) \leq c$
- The solution of this problem is $x^*(\theta)$, and the optimal value of $f$ is $f(x^*(\theta), \theta)$
- Now define the value function $V$ : **R** $\to$ **R** as $V(\theta) \equiv f(x^*(\theta), \theta) = \max_x f(x, \theta)$ s.t. $g(x, \theta) \leq c$
- Question: evaluate how $V$ changes with $\theta$.
- Example: Utility ($U(c_1, c_2)$) maximization. Income and prices ($I, p_1, p_2$) are exogenous. Thus optimal demands are $c_1^*(I, p_1, p_2), c_2^*(I, p_1, p_2)$ and $\lambda^*(I, p_1, p_2)$.
- $V(I) \equiv U(c_1^*(I, \cdot), c_2^*(I, \cdot))$.

# The envelope theorem (2)

## Evaluating $V$

- Suppose in particular that we want to assess $V'(\theta)$.
- From K-T conditions, we know that $(x^*, \lambda^*)$ must satisfy the complementary slackness condition: $\lambda^*[c - g(x^*, \theta)] = 0$
- Thus, $V(\theta) \equiv f(x^*, \theta) = f(x^*, \theta) + \lambda^*[c - g(x^*, \theta)]$
- Differentiate both sides wrt $\theta:\ V'(\theta) = \frac{\partial f}{\partial \theta} - \lambda^* \frac{\partial g}{\partial \theta}$
- BUT we know that $x^*(\theta)$ and $\lambda^*(\theta)$. Then, that derivative may be wrong when ignoring the dependence of $x^*$ and $\lambda^*$ on $\theta$.
- HOWEVER, the envelope theorem tells us that the expression of $V'(\theta)$ is in fact correct, i.e.
- The envelope theorem tells us that when computing $V'(\theta)$ we can ignore the dependence of $x^*$ and $\lambda^*$ on $\theta$.

# The envelope theorem - Intuition

## Unconstrained problem $\max_x f(x,\theta)$

- Let $x^*(\theta)$ be the solution, and define $V(\theta) \equiv f(x^*(\theta),\theta)$.
- Differentiate both sides wrt $\theta$ : $V'(\theta) = \frac{\partial f}{\partial x^*}\frac{dx^*}{d\theta} + \frac{\partial f}{\partial \theta}$
- but as $x^*(\theta)$ is a critical value of $f$, it must be that $\frac{\partial f}{\partial x^*} = 0$
- Therefore, $V'(\theta) = \frac{\partial f}{\partial \theta}$

# The envelope theorem - Intuition (2)

## Constrained problem $\max_x f(x,\theta)$ s.t. $g(x,\theta) \le c$

- From K-T conditions, we know that $(x^*(\theta), \lambda^*(\theta))$ must satisfy the complementary slackness condition:
  $\lambda^*(\theta)[c - g(x^*(\theta),\theta)] = 0$
- Thus, $V(\theta) = f(x^*(\theta),\theta) + \lambda^*(\theta)[c - g(x^*(\theta),\theta)]$
- Differentiate both sides wrt $\theta$ :
  $V'(\theta) =$
  $\frac{\partial f}{\partial x^*}\frac{dx^*}{d\theta} + \frac{\partial f}{\partial \theta} + \frac{d\lambda^*(\theta)}{d\theta}[c - g(x^*(\theta),\theta)] - \lambda^*(\theta)[\frac{\partial g}{\partial x^*}\frac{dx^*}{d\theta} + \frac{\partial g}{\partial \theta}] =$
  $\frac{dx^*}{d\theta}\left[\frac{\partial f}{\partial x^*} - \lambda^*(\theta)\frac{\partial g}{\partial x^*}\right] + \frac{\partial f}{\partial \theta} + \frac{d\lambda^*(\theta)}{d\theta}\Big[c - g(x^*(\theta),\theta)\Big] - \lambda^*(\theta)\frac{\partial g}{\partial \theta}$
- BUT $(x^*(\theta), \lambda^*(\theta)$ is a critical point of the lagrangean function $L(x,\lambda) = f(x,\cdot) + \lambda[c - g(x,\cdot)]$. Therefore,
- $\frac{\partial L}{\partial x^*} = \frac{\partial f}{\partial x^*} - \lambda^*\frac{\partial g}{\partial x^*} = 0$, and $\frac{\partial L}{\partial \lambda^*} = c - g(x^*,\theta) = 0$. Thus,
- $V'(\theta) = \frac{\partial f}{\partial \theta} - \lambda^*(\theta)\frac{\partial g}{\partial \theta}$

# The envelope theorem

- Let $f(x,\theta)$ and $g(x,\theta)$ be contiunuously differentiable functions.
- For any given $\theta$, $x^*(\theta)$ maximizes $f(x,\theta)$ s.t. $g(x,\theta) \leq c$.
- Let $\lambda^*(\theta)$ be the value of the associated lagrange multiplier.
- Suppose $x^*(\theta)$ and $\lambda^*(\theta)$ be continuously differentuiable.
- Suppose that the constraint qualification, $g(x^*(\theta),\theta) \neq 0$ holds $\forall\theta$.
- Then, the maximum value function defined by
  $V(\theta) = \max_x f(x,\theta)$ s.t. $g(x,\theta) \leq c$
  satisfies
  $V'(\theta) = \frac{\partial f(x^*(\theta),\theta)}{\partial\theta} - \lambda^*(\theta)\frac{\partial g(x^*(\theta),\theta)}{\partial\theta}$

# The envelope theorem - Proof

- K-T theorem says that for any $\theta, x^*(\theta)$ and $\lambda^*(\theta)$ satisfy $\frac{\partial L(x^*(\theta),\lambda^*(\theta))}{\partial x} = \frac{\partial f(x^*(\theta),\theta)}{\partial x} - \lambda^*(\theta)\frac{\partial g(x^*(\theta),\theta)}{\partial x} = 0$, [1] and $\lambda^*[c - g(x^*(\theta),\theta)] = 0$ [2]
- Define $V(\theta) = f(x^*(\theta),\theta) + \lambda^*(\theta)[c - g(x^*(\theta),\theta)]$
- Differentiate both sides wrt $\theta$ : $V'(\theta) = \frac{\partial f}{\partial x^*}\frac{dx^*}{d\theta} + \frac{\partial f}{\partial \theta} + \frac{d\lambda^*(\theta)}{d\theta}[c - g(x^*(\theta),\theta)] - \lambda^*(\theta)[\frac{\partial g}{\partial x^*}\frac{dx^*}{d\theta} + \frac{\partial g}{\partial \theta}]$
- Rewrite as $V'(\theta) =$ $\left[\frac{\partial f}{\partial x^*} - \lambda^*(\theta)\frac{\partial g}{\partial x^*}\right]\frac{dx^*}{d\theta} + \frac{\partial f}{\partial \theta} - \lambda^*(\theta)\frac{\partial g}{\partial \theta} + \frac{d\lambda^*(\theta)}{d\theta}[c - g(x^*(\theta),\theta)]$
- The first term is zero from the FOC [1]
- If the constraint binds the last term is also zero.

# The envelope theorem - Proof (cont'd)

- If the constraint does not bind, $\lambda^*(\theta) = 0$
- The continuity of $g$ and $x^*$ means that if the constraint does not bind at $\theta, \exists \varepsilon^* > 0$ such that the constraint does not bind for $\theta + \varepsilon$ with $|\varepsilon| < \varepsilon^*$.
- FOC [2] implies that $\lambda^*(\theta + \varepsilon) = 0, \forall |\varepsilon| < \varepsilon^*$.
- From the definition of derivative $\frac{d\lambda^*(\theta)}{d\theta} = \lim_{\varepsilon \to 0} \frac{\lambda^*(\theta+\varepsilon) - \lambda^*(\theta)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{0}{\varepsilon} = 0$, hence the last term is again zero.
- Therefore, it follows that $V'(\theta) = \frac{\partial f(x^*(\theta),\theta)}{\partial \theta} - \lambda^*(\theta) \frac{\partial g(x^*(\theta),\theta)}{\partial \theta}$

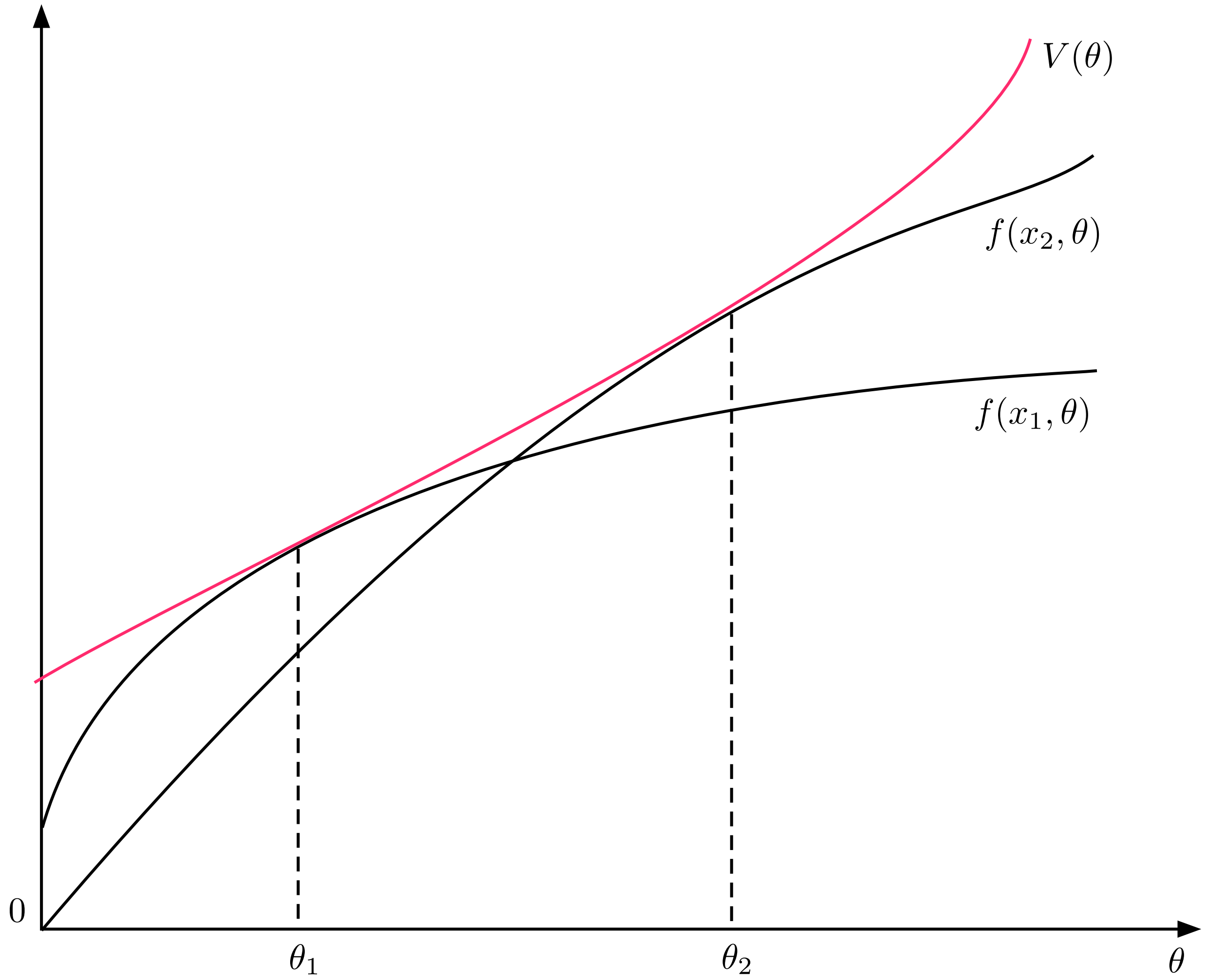
[See Ireland, 2010]

# The envelope theorem - Geometry

## Unconstrained problem $\max_x f(x,\theta)$

- Let $x^*(\theta)$ be the solution, and define $V(\theta) = \max_x f(x,\theta)$.
- Let $\theta_1$ be a particular value of $\theta$ and let $x_1 = x^*(\theta_1)$.
- Think of $f(x_1,\theta)$ a function of $\theta$ holding $x_1$ fixed.
- Similarly consider $\theta_2 > \theta_1$ and $x_2 = x^*(\theta_2)$
- Think of $f(x_2,\theta)$ a function of $\theta$ holding $x_2$ fixed.
- For $\theta = \theta_1$ because $x_1$ maximizes $f(x,\theta_1)$ it follows that $V(\theta_1) = f(x_1,\theta_1) > f(x_2,\theta_1)$
- For $\theta = \theta_2$ because $x_2$ maximizes $f(x,\theta_2)$ it follows that $V(\theta_2) = f(x_2,\theta_2) > f(x_1,\theta_2)$
- Graphically, at $\theta_1$, $V(\theta) = f(x_1,\theta)$ which lies above $f(x_2,\theta)$.
- Graphically, at $\theta_2$, $V(\theta) = f(x_2,\theta)$ which lies above $f(x_1,\theta)$.

# The envelope theorem - Geometry (2)



# The envelope theorem - Geometry (3)

- Repeat the argument for more values of $\theta_i$, $i = 1, 2, 3, \ldots$
- $V(\theta)$ is tangent to each $f(x_i, \theta)$ at $\theta_i$, i.e. $V'(\theta) = \frac{\partial f(x^*(\theta),\theta)}{\partial \theta}$
- This is the same analytical result obtained.
- For the constrained maximization problem $\max_x f(x, \theta)$ s.t. $g(x, \theta) \leq c$, define the lagrangean function $L(x, \lambda, \theta) = f(x, \theta) + \lambda(c - g(x, \theta))$
- Define $V(\theta) = \max_x L(x, \lambda, \theta)$ and repeat the argument above wrt $L$
- Again $V'(\theta) = \frac{\partial L}{\partial \theta} = \frac{\partial f(x^*(\theta),\theta)}{\partial \theta} - \lambda^*(\theta) \frac{\partial g(x^*(\theta),\theta)}{\partial \theta}$
- The $V$ function is tangent from above to all the $L$ functions associated to the values $\theta_i$

# Applications

- Consumer Theory (Utility max)
- Producer Theory
  - Profit maximization
  - Cost minimization
  - Revenue max under profit constraint
  - Peak-Load pricing
  - Regulatory constraints: rate-of-return, environmental, ...
- Welfare economics (Pareto optimal solutions)
- Human capital investment
- Non-Linear Least-Square estimation
- ... and many, many more