Chapter 6 Basics of Set-Constrained and Unconstrained Optimization

An Introduction to Optimization

Spring, 2014

Wei-Ta Chu

Introduction

- Consider the optimization problem

$$\begin{aligned}&\text{minimize } f(\boldsymbol{x})\\&\text{subject to } \boldsymbol{x} \in \Omega\end{aligned}$$

- The function $f : R^n \to R$ that we wish to minimize is a real-valued function called the ***objective function*** or ***cost function***. The vector $\boldsymbol{x} = [x_1, x_2, ..., x_n]^T \in R^n$. The variables $x_1, x_2, ..., x_n$ are often referred to as ***decision variables***. The set Ω is a subset of R^n called the ***constraint set*** or ***feasible set***.
- Finding the "best" vector $\boldsymbol{x}$ over all possible vectors in Ω. This vector is called the ***minimizer*** of f over Ω. It is possible that there may be many minimizers.

Introduction

- There are also optimization problems that require maximization of the objective function, in which case seek ***maximizers***. Minimizers and maximizers are also called ***extremizers***.
- Maximization problems can be represented equivalently in the minimization form because maximizing f is equivalent to minimizing $-f$.
- This problem is a general form of a ***constrained optimization problem***. If $\Omega = R^n$, then we refer to the problem as an ***unconstrained optimization problem***.

Introduction

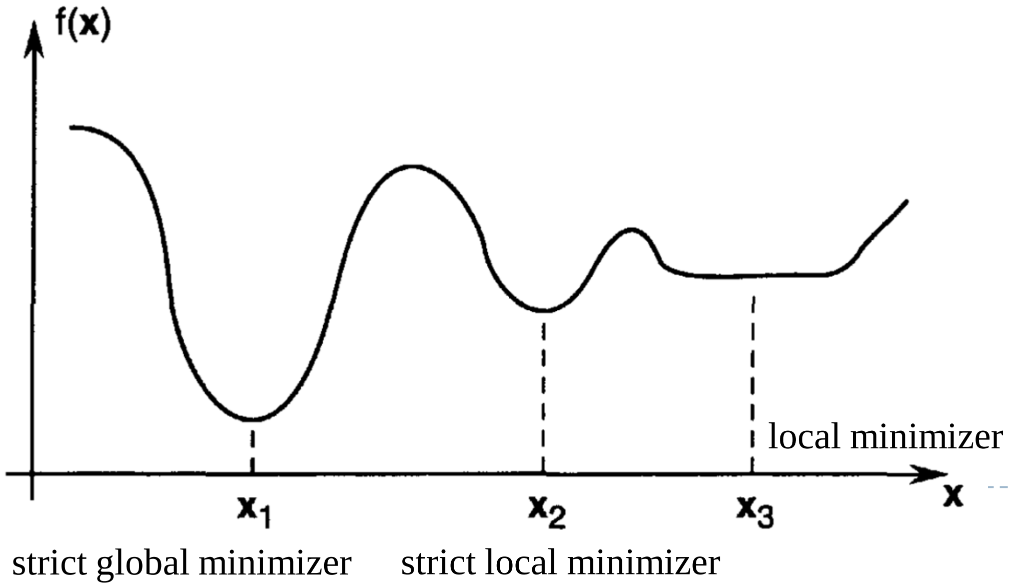
- The constraint $\boldsymbol{x} \in \Omega$ is called a ***set constraint***.
- Often, the constraint set Ω takes the form

$$\Omega = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$$

  where $\boldsymbol{h}$ and $\boldsymbol{g}$ are given functions. We refer to such constraints as ***functional constraints***.

Global and Local Minimizers

- Definition 6.1: Suppose that $f: R^n \to R$ is a real-valued function defined on some set $\Omega \subset R^n$. A point $\boldsymbol{x}^* \in \Omega$ is a ***local minimizer*** of f over Ω if there exists $\epsilon > 0$ such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon$ and $\boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\}$. A point $\boldsymbol{x}^* \in \Omega$ is a ***global minimizer*** of f over Ω if $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\}$.
- If we replace $\geq$ by $>$, then we have a ***strict local minimizer*** and a ***strict global minimizer***, respectively.



Global and Local Minimizers

- If $\boldsymbol{x}^*$ is a global minimizer of f over Ω, we write $f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$ and $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$.
- If the minimization is unconstrained, we simply write $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$ or $\boldsymbol{x}^* = \arg\min f(\boldsymbol{x})$
- Example: if $f : R \to R$ is given by $f(x) = (x+1)^2 + 3$, then $\arg\min f(x) = -1$
- Strictly speaking, an optimization problem is solved only when a global minimizer is found. However, global minimizers are, in general, difficult to find. Therefore, in practice, we often have to be satisfied with finding local minimizers.

Conditions for Local Minimizers

- Given a function $f : R^n \to R$, recall that the first-order derivative of f , denoted by Df , is

$$Df \triangleq \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right]$$

- The gradient $\bigtriangledown f$ is just the transpose of Df; that is, $\bigtriangledown f = (Df)^T$
- The second derivative of f (also called *Hessian* of f) is

$$\boldsymbol{F}(\boldsymbol{x}) \triangleq D^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{x}) \end{bmatrix}$$

Conditions for Local Minimizers

- Example: Let

$$f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$$

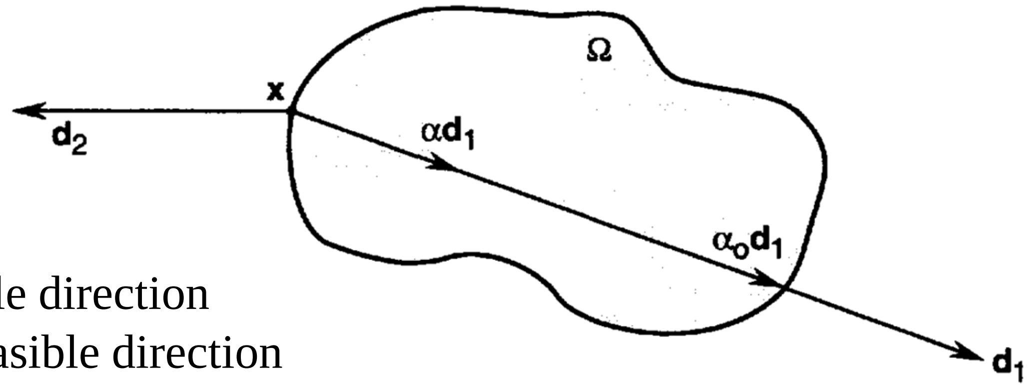
$$Df(\boldsymbol{x}) = (\triangledown f(\boldsymbol{x}))^T = \left[\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x})\right] = [5 + x_2 - 2x_1, 8 + x_1 - 4x_2]$$

$$\boldsymbol{F}(\boldsymbol{x}) \triangleq D^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}$$

Feasible Direction

- Definition 6.2: A vector $\boldsymbol{d} \in R^n, \boldsymbol{d} \neq \mathbf{0}$ is a ***feasible direction*** at $\boldsymbol{x} \in \Omega$ if there exists $\alpha_0 > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$
- The ***directional derivative of*** f ***in the direction*** $\boldsymbol{d}$, denoted $\frac{\partial f}{\partial \boldsymbol{d}}$, is the real-valued function defined by

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \lim_{\alpha \to 0} \frac{f(\boldsymbol{x} + \alpha \boldsymbol{d}) - f(\boldsymbol{x})}{\alpha}$$



$\boldsymbol{d}_1$ is a feasible direction
$\boldsymbol{d}_2$ is a not feasible direction

Directional Derivative

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \lim_{\alpha \to 0} \frac{f(\boldsymbol{x} + \alpha \boldsymbol{d}) - f(\boldsymbol{x})}{\alpha}$$

- If $\|\boldsymbol{d}\| = 1$, then $\frac{\partial f}{\partial \boldsymbol{d}}$ is the rate of increase of f at $\boldsymbol{x}$ in the direction $\boldsymbol{d}$.
- Suppose that $\boldsymbol{x}$ and $\boldsymbol{d}$ are given. Then $f(\boldsymbol{x} + \alpha \boldsymbol{d})$ is a function of α, and

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \frac{\partial}{\partial \alpha} f(\boldsymbol{x} + \alpha \boldsymbol{d}) \Big|_{\alpha=0}$$

- Applying the chain rule yields

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \frac{\partial}{\partial \alpha} f(\boldsymbol{x} + \alpha \boldsymbol{d}) \Big|_{\alpha=0} = \bigtriangledown f(\boldsymbol{x})^T \boldsymbol{d} = \langle \bigtriangledown f(\boldsymbol{x}), \boldsymbol{d} \rangle = \boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x})$$

- In summary, if $\boldsymbol{d}$ is a unit vector, then $\langle \bigtriangledown f(\boldsymbol{x}), \boldsymbol{d} \rangle$ is the rate of increase of f at the point $\boldsymbol{x}$ in the direction $\boldsymbol{d}$.

$$\boxed{\frac{d}{dx} f(g(x)) = f'(g(x)) g'(x)}$$

Example

- Define $f : R^3 \to R$ by $f(\boldsymbol{x}) = x_1x_2x_3$ and let $\boldsymbol{d} = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right]^T$

  The directional derivative of f in the direction $\boldsymbol{d}$

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \bigtriangledown f(\boldsymbol{x})^T \boldsymbol{d} = [x_2x_3, x_1x_3, x_1x_2] \begin{bmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix} = \frac{x_2x_3 + x_1x_3 + \sqrt{2}x_1x_2}{2}$$

- Note that because $\|\boldsymbol{d}\| = 1$, the above is also the rate of increase of f at $\boldsymbol{x}$ in the direction $\boldsymbol{d}$

Theorem 6.1 First-Order Necessary Condition (FONC)

- Let Ω be a subset of R^n and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, we have

$$\boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}^*) \geq 0$$

- Proof: Define $\boldsymbol{x}(\alpha) = \boldsymbol{x}^* + \alpha \boldsymbol{d} \in \Omega$. Note that $\boldsymbol{x}(0) = \boldsymbol{x}^*$. Define the composite function

$$\phi(\alpha) = f(\boldsymbol{x}(\alpha))$$

Then, by Taylor's theorem

$$f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) - f(\boldsymbol{x}^*) = \phi(\alpha) - \phi(0) = \phi'(0)\alpha + o(\alpha) = \alpha \boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}(0)) + o(\alpha)$$

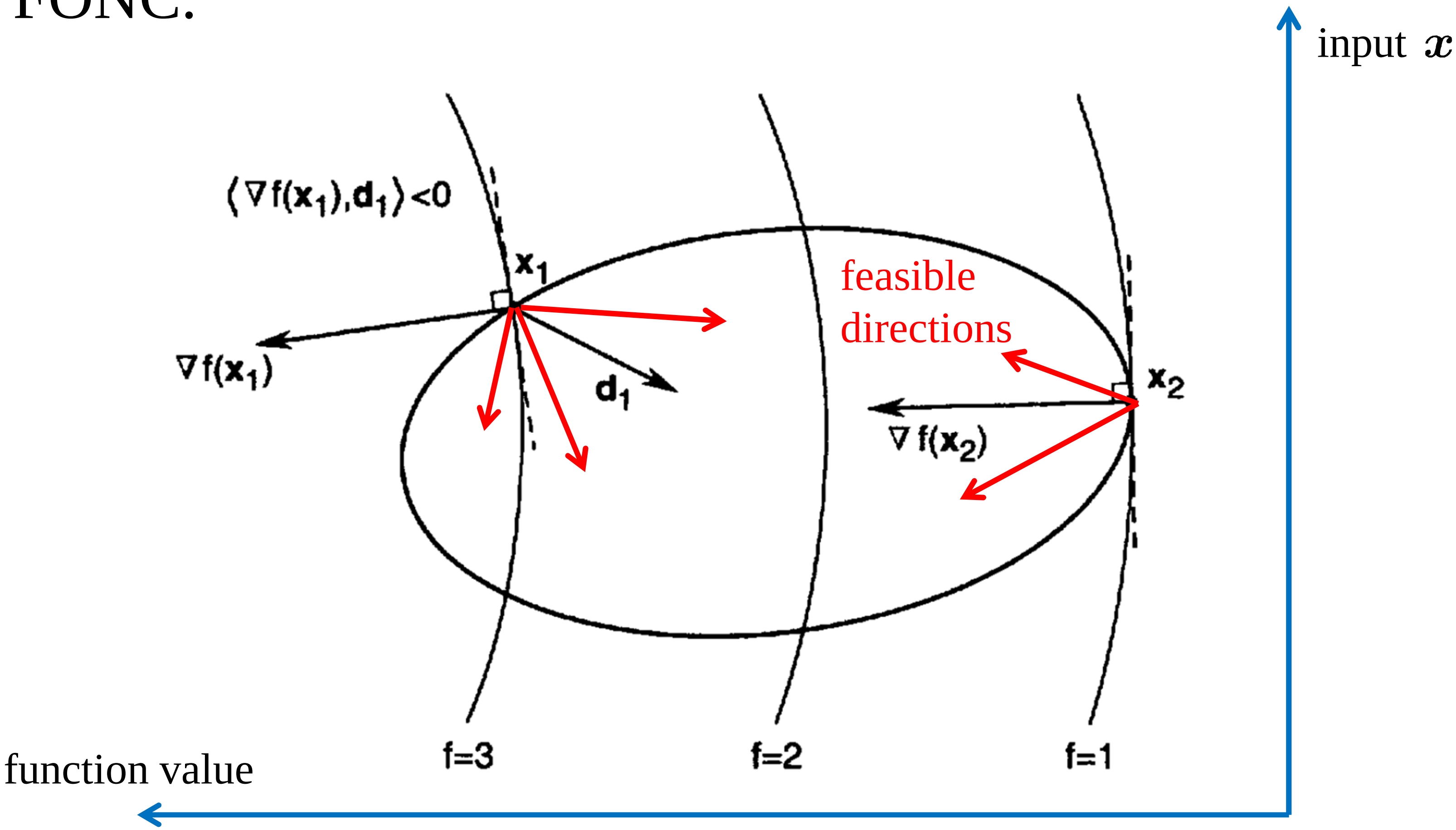
where $\alpha \geq 0$. Thus, if $\phi(\alpha) \geq \phi(0)$, that is, $f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) \geq f(\boldsymbol{x}^*)$ for sufficiently small values of $\alpha > 0$ ($\boldsymbol{x}^*$ is a local minimizer), then we have to have $\boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}^*) \geq 0$.

$$\boxed{f(b) = f(a) + \frac{h}{1!} f^{(1)}(a) + \frac{h^2}{2!} f^{(2)}(a) + \cdots + \frac{h^{m-1}}{(m-1)!} f^{(m-1)}(a) + \frac{h^m}{m!} f^{(m)}(a) + o(h^m)}$$

$\phi(\alpha)$ $\phi(0)$ $h = b - a = \alpha$

First-Order Necessary Condition (FONC)

- x_1 does not satisfy the FONC, whereas x_2 satisfies the FONC.



First-Order Necessary Condition (FONC)

- An alternative way to express the FONC is $\dfrac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}^*) \geq 0$ for all feasible directions $\boldsymbol{d}$.
- In other words, if $\boldsymbol{x}^*$ is a local minimizer, then the rate of increase of f at $\boldsymbol{x}^*$ in any feasible direction $\boldsymbol{d}$ in Ω is nonnegative.
- Alternative proof: For any feasible direction, there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha})$,
$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}^* + \alpha \boldsymbol{d})$$
Hence, for all $\alpha \in (0, \bar{\alpha})$
$$\frac{f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) - f(\boldsymbol{x}^*)}{\alpha} \geq 0$$
Taking the limit $\alpha \to 0$, we conclude $\dfrac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}^*) \geq 0$

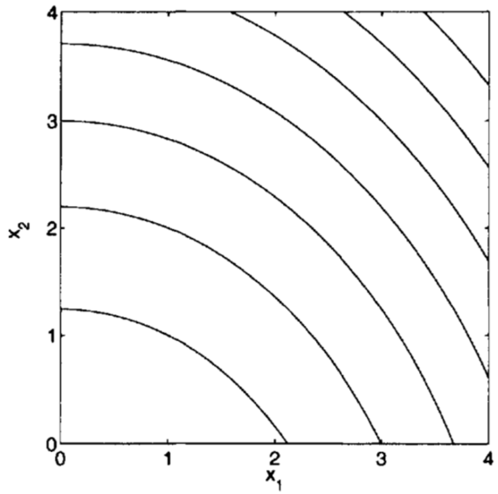
Corollary 6.1 Interior Case

- Let Ω be a subset of R^n and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω and if $\boldsymbol{x}^*$ is an interior point of Ω, then
$$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$$
- Proof: Suppose that f has a local minimizer $\boldsymbol{x}^*$ that is an interior point of Ω. The set of feasible directions at $\boldsymbol{x}^*$ is the whole of R^n. Thus, for any $\boldsymbol{d} \in R^n$, $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) \geq 0$ and $-\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) \geq 0$. Hence, $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) = 0$ for all $\boldsymbol{d} \in R^n$, which implies that $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.

Example

- Consider the problem
$$\begin{aligned}&\text{minimize } x_1^2 + 0.5x_2^2 + 3x_2 + 4.5\\&\text{subject to } x_1, x_2 \geq 0\end{aligned}$$
- Problem 1: Is the first-order necessary condition for a local minimizer satisfied at $\boldsymbol{x} = [1,3]^T$?
- Solution: At $\boldsymbol{x} = [1,3]^T$, we have $\bigtriangledown f(\boldsymbol{x}) = [2x_1, x_2+3]^T = [2,6]^T$ The point is an interior point of $\Omega = \{\boldsymbol{x} : x_1 \geq 0, x_2 \geq 0\}$. Hence, the FONC requires that $\bigtriangledown f(\boldsymbol{x}) = \mathbf{0}$. Therefore, the point $\boldsymbol{x} = [1,3]^T$ doest not satisfy the FONC for a local minimizer.

Level sets of this function



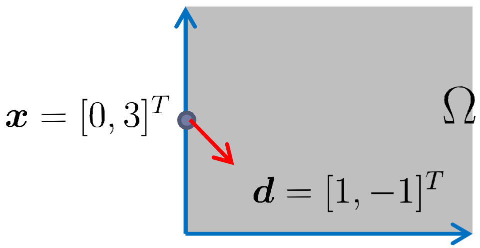
$\nabla f(\boldsymbol{x}) = [2x_1, x_2 + 3]^T$

Example

- Consider the problem $\begin{aligned} &\text{minimize } x_1^2 + 0.5x_2^2 + 3x_2 + 4.5 \\ &\text{subject to } x_1, x_2 \geq 0 \end{aligned}$

On the boundary of Ω

- Problem 2: Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [0, 3]^T$
- Solution: At $\boldsymbol{x} = [0, 3]^T$, we have $\nabla f(\boldsymbol{x}) = [0, 6]^T$, and hence $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = 6d_2$, where $\boldsymbol{d} = [d_1, d_2]^T$. For $\boldsymbol{d}$ to be feasible at $\boldsymbol{x}$, we need $d_1 \geq 0$ and d_2 can take an arbitrary value in *R*. The point $\boldsymbol{x} = [0, 3]^T$ does not satisfy the FONC for a minimizer because d_2 is allowed to be less than zero.
- For example, $\boldsymbol{d} = [1, -1]^T$ is a feasible direction, but $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = -6 < 0$



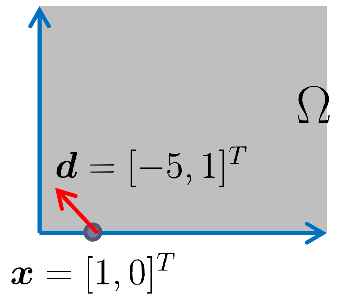
$$\nabla f(\boldsymbol{x}) = [2x_1, x_2+3]^T$$

Example

- Consider the problem $\text{minimize } x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$
  $\text{subject to } x_1, x_2 \geq 0$

On the boundary of Ω

- Problem 3: Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [1,0]^T$
- Solution: At $\boldsymbol{x} = [1,0]^T$, we have $\nabla f(\boldsymbol{x}) = [2,3]^T$, and hence $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = 2d_1 + 3d_2$. For $\boldsymbol{d}$ to be feasible, we need $d_2 \geq 0$ and d_1 can take an arbitrary value in *R*. For example, $\boldsymbol{d} = [-5,1]^T$ is a feasible solution. But $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = -7 < 0$. Thus, $\boldsymbol{x} = [1,0]^T$ does not satisfy the FONC for a local minimizer.



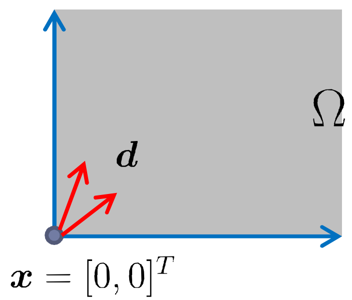
$$\nabla f(\boldsymbol{x}) = [2x_1, x_2 + 3]^T$$

Example

- Consider the problem $\text{minimize } x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$
  $\text{subject to } x_1, x_2 \geq 0$

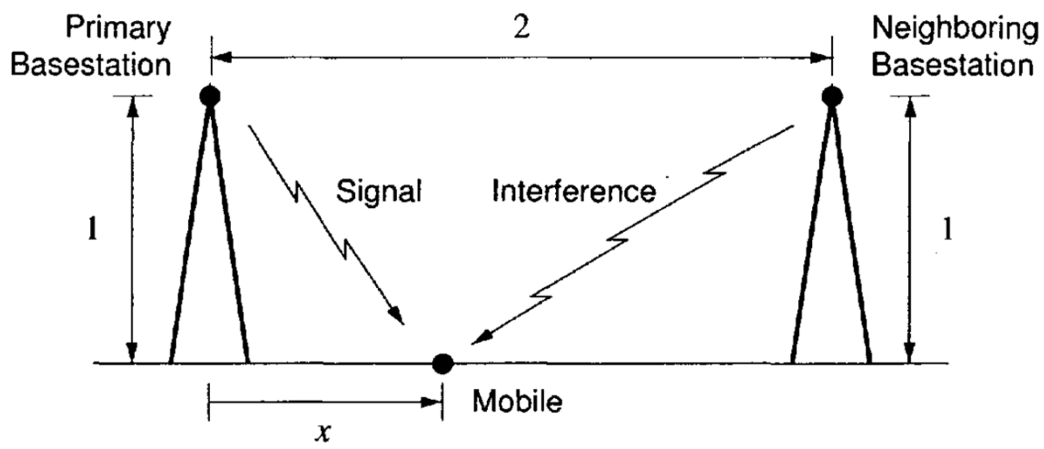
  On the boundary of Ω

- Problem 4: Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [0, 0]^T$
- Solution: At $\boldsymbol{x} = [0, 0]^T$, we have $\nabla f(\boldsymbol{x}) = [0, 3]^T$, and hence $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = 3d_2$. For $\boldsymbol{d}$ to be feasible, we need $d_2 \geq 0$ and $d_1 \geq 0$. Hence, $\boldsymbol{x} = [0, 0]^T$ satisfies the FONC for a local minimizer.

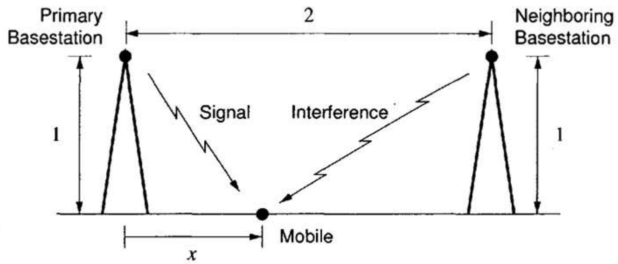


Example

- There are two base station antennas, one for the primary base station and another for the neighboring base station. Both antennas are transmitting signals to the mobile user, at equal power.
- The power of the received signal is reciprocal of the squared distance from the associated antenna.
- Find the position of the mobile that maximizes the signal-to-interference ratio.



Example



- The squared distance from the mobile to the primary antenna is $1+x^2$, while the squared distance from the mobile to the neighboring antenna is $1+(2-x)^2$. Therefore, the signal-to-inference ratio is

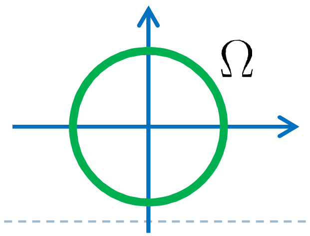
$$f(x)=\frac{1+(2-x)^2}{1+x^2}$$

$$f'(x)=\frac{-2(2-x)(1+x^2)-2x(1+(2-x)^2)}{(1+x^2)^2}=\frac{4(x^2-2x-1)}{(1+x^2)^2}$$

- By the FONC, at the optimal position x^* we have $f'(x^*)=0$. Hence, either $x^*=1-\sqrt{2}$ or $x^*=1+\sqrt{2}$. Evaluating the objective function at these two candidate points, it's easy to see that $x^*=1-\sqrt{2}$ is the optimal solution.

$$\frac{d}{dx}\left(\frac{f}{g}\right)=\frac{f'g-fg'}{g^2}$$

Example



- Consider the set-constrained problem where $\Omega = \{[x_1, x_2]^T : x_1^2 + x_2^2 = 1\}$

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega \end{array}$$

- Problem 1: Consider a point $\boldsymbol{x}^* \in \Omega$. Specify all feasible directions at $\boldsymbol{x}^*$.
- Solution: There are no feasible directions at any $\boldsymbol{x}^*$.

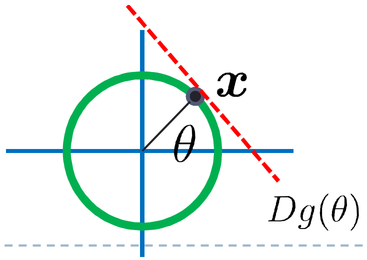
- Problem 2: Which points in Ω satisfy the FONC for this set-constrained problem?
- Solution: Because of the solution for Problem 1, all points in Ω satisfy the FONC for this set-constrained problem.

$$\text{minimize } f(\boldsymbol{x})$$
$$\text{subject to } x \in \Omega$$
$$\Omega = \{[x_1, x_2]^T : x_1^2 + x_2^2 = 1\}$$

Example

- Problem 3: Based on Problem 2, is the FONC for this set-constrained problem useful for eliminating local-minimizer candidates?
- Solution: No, the FONC for this set-constrained problem is not useful for eliminating local-minimizer candidates.

Example



$$\begin{aligned}&\text{minimize } f(\boldsymbol{x})\\&\text{subject to } x \in \Omega\\&\Omega = \{[x_1, x_2]^T : x_1^2 + x_2^2 = 1\}\end{aligned}$$

- Problem 4: Suppose that we use polar coordinates to parameterize points $\boldsymbol{x} \in \Omega$ in terms of a single parameter θ : $x_1 = \cos\theta$ and $x_2 = \sin\theta$. Now use the FONC for unconstrained problems (with respect to θ) to derive a necessary condition of this sort: if $\boldsymbol{x}^* \in \Omega$ is a local minimizer, then $\boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}^*) = 0$ for all $\boldsymbol{d}$ satisfying a "certain condition." Specify what this certain condition is.
- Solution:
  - Write $h(\theta) = f(g(\theta))$, where $g : R \to R^2$ is given by the equations relating θ to $\boldsymbol{x} = [x_1, x_2]^T$. Note that $Dg(\theta) = [-\sin\theta, \cos\theta]^T$. Hence,
$$h'(\theta) = Df(g(\theta))Dg(\theta) = Dg(\theta)^T \bigtriangledown f(g(\theta))$$
  - Notice that $Dg(\theta)$ is tangent to Ω at $\boldsymbol{x} = g(\theta)$. Alternatively, we could say that $Dg(\theta)$ is orthogonal to $\boldsymbol{x} = g(\theta)$.

Example

$$h'(\theta) = Df(g(\theta))Dg(\theta) = Dg(\theta)^T \bigtriangledown f(g(\theta))$$

- Solution:
  - Suppose that $\boldsymbol{x}^* \in \Omega$ is a local minimizer. Write $\boldsymbol{x}^* = g(\theta^*)$. Then θ^* is an unconstrained minimizer of h. By the FONC for unconstrained problems, $h'(\theta^*) = 0$, which implies that $\boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}^*) = 0$ for all $\boldsymbol{d}$ tangent to Ω at $\boldsymbol{x}^*$ (or, alternatively, for all $\boldsymbol{d}$ orthogonal to $\boldsymbol{x}^*$)

Same as the Corollary 6.1 Interior Case

Theorem 6.2 Second-Order Necessary Condition (SONC)

- Let $\Omega \subset R^n$, $f \in \mathcal{C}^2$ a function on Ω , $\boldsymbol{x}^*$ a local minimizer of f over Ω, and $\boldsymbol{d}$ a feasible direction at $\boldsymbol{x}^*$. If $\boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}^*) = 0$, then

  $$\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$$

  Corollary 6.1 Interior Case

  where $\boldsymbol{F}$ is the Hessian of f .
- Proof: We prove the result by contradiction. Suppose that there is a feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$ such that $\boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}^*) = 0$ and $\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} < 0$. Let $\boldsymbol{x}(\alpha) = \boldsymbol{x}^* + \alpha\boldsymbol{d}$ and define the composite function $\phi(\alpha) = f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) = f(\boldsymbol{x}(\alpha))$.
  By Taylor's theorem,

  $$\phi(\alpha) = \phi(0) + \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2)$$

  where by assumption, $\phi'(0) = \boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}^*) = 0$ and $\phi''(0) = \boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} < 0$.

Theorem 6.2 Second-Order Necessary Condition (SONC)

$$\phi'(0) = \boldsymbol{d}^T \bigtriangledown f(\boldsymbol{x}^*) = 0 \qquad \phi''(0) = \boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} < 0$$

- For sufficiently small α ,

$$\phi(\alpha) - \phi(0) = \phi''(0)\tfrac{\alpha^2}{2} + o(\alpha^2) < 0$$

that is,

$$f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) < f(\boldsymbol{x}^*)$$

which contradicts the assumption that $\boldsymbol{x}^*$ is a local minimizer. Thus,

$$\phi''(0) = \boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$$

$$\boxed{\begin{aligned} \phi(\alpha) &= f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) = f(\boldsymbol{x}(\alpha)) \\ \phi(0) &= f(\boldsymbol{x}^* + 0\boldsymbol{d}) = f(\boldsymbol{x}^*) \end{aligned}}$$

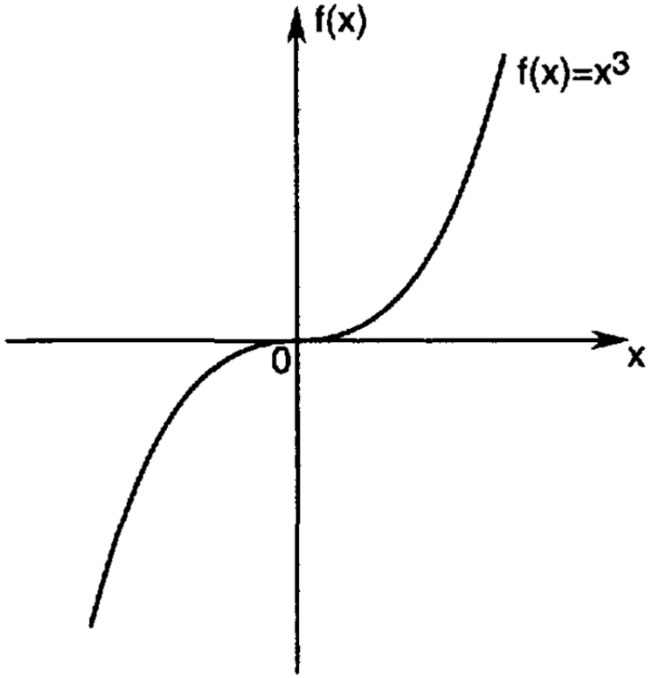
Corollary 6.2 Interior Case

- Let $\boldsymbol{x}^*$ be an interior point of $\Omega \subset R^n$. If $\boldsymbol{x}^*$ is a local minimizer of $f : \Omega \to R$, $f \in \mathcal{C}^2$, then $\bigtriangledown f(\boldsymbol{x}^*) = \mathbf{0}$, and $\boldsymbol{F}(\boldsymbol{x}^*)$ is positive semidefinite $(\boldsymbol{F}(\boldsymbol{x}^*) \geq 0)$; that is, for all $\boldsymbol{d} \in R^n$, $\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$
- Proof: If $\boldsymbol{x}^*$ is an interior point, then all directions are feasible. The result then follows from Corollary 6.1 and Theorem 6.2.

Example

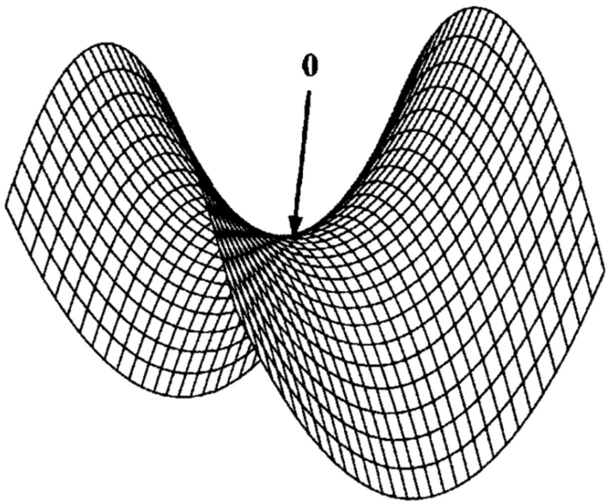
- Consider a function of one variable $f(x) = x^3, f : R \to R.$
- Because $f'(0) = 0$ and $f''(0) = 0$, the point $x = 0$ satisfies both the FONC and SONC.
- However, $x = 0$ is not a minimizer.

They are necessary conditions, but are not sufficient conditions.



Example

- Consider a function $f : R^2 \to R, f(\boldsymbol{x}) = x_1^2 - x_2^2$. The FONC requires that $\nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^T = \mathbf{0}$. Thus, $\boldsymbol{x} = [0, 0]^T$ satisfies the FONC.
- The Hessian matrix of f is $\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$
- The Hessian matrix is indefinite; that is, for some $\boldsymbol{d}_1 \in R^2$ we have $\boldsymbol{d}_1^T \boldsymbol{F} \boldsymbol{d}_1 > 0$ (e.g., $\boldsymbol{d}_1 = [1, 0]^T$); and, for some $\boldsymbol{d}_2$, we have $\boldsymbol{d}_2^T \boldsymbol{F} \boldsymbol{d}_2 < 0$ (e.g., $\boldsymbol{d}_2 = [0, 1]^T$). Thus, $\boldsymbol{x} = [0, 0]^T$ does not satisfy the SONC, and hence it is not a minimizer.



Theorem 6.3 Second-Order Sufficient Condition (SOSC), Interior Case

- Let $f \in \mathcal{C}^2$ be defined on a region in which $\boldsymbol{x}^*$ is an interior point. Suppose that $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ and $\boldsymbol{F}(\boldsymbol{x}^*) > 0$. Then, $\boldsymbol{x}^*$ is a strict local minimizer of f .
- Proof: Because $f \in \mathcal{C}^2$, we have $\boldsymbol{F}(\boldsymbol{x}^*) = \boldsymbol{F}^T(\boldsymbol{x}^*)$. Using assumption 2 and Rayleigh's inequality it follows that if $\boldsymbol{d} \neq \boldsymbol{0}$ then $0 < \lambda_{min}(\boldsymbol{F}(\boldsymbol{x}^*))\|\boldsymbol{d}\|^2 \leq \boldsymbol{d}^T\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d}$. By the Taylor's theorem and assumption 1,

$$f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) - f(\boldsymbol{x}^*) = \tfrac{1}{2}\boldsymbol{d}^T\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} + o(\|\boldsymbol{d}\|^2) \geq \frac{\lambda_{min}(\boldsymbol{F}(\boldsymbol{x}^*))}{2}\|\boldsymbol{d}\|^2 + o(\|\boldsymbol{d}\|^2) > 0$$

(p. 27)

Hence, for all $\boldsymbol{d}$ such that $\|\boldsymbol{d}\|$ is sufficiently small,

$$f(\boldsymbol{x}^* + \boldsymbol{d}) > f(\boldsymbol{x}^*)$$

which completes the proof.

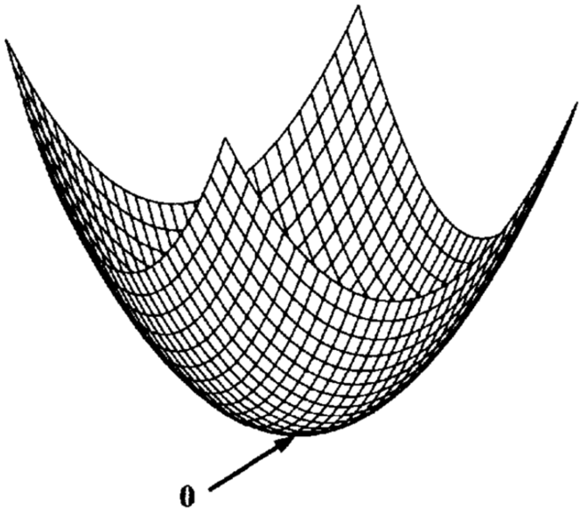
$$\boxed{\lambda_{min}(\boldsymbol{P})\|\boldsymbol{x}\|^2 \leq \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} \leq \lambda_{max}(\boldsymbol{P})\|\boldsymbol{x}\|^2}$$

Example

- Let $f(\boldsymbol{x}) = x_1^2 + x_2^2$. We have $\nabla f(\boldsymbol{x}) = [2x_1, 2x_2]^T = \mathbf{0}$ if and only if $\boldsymbol{x} = [0, 0]^T$. For all $\boldsymbol{x} \in R^2$, we have

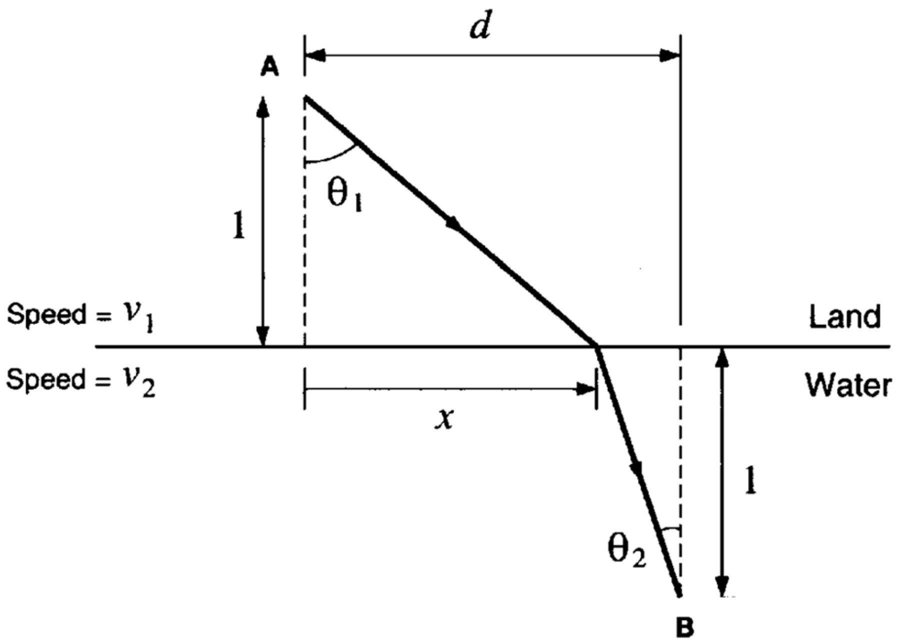
$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0$$

- The point $\boldsymbol{x} = [0, 0]^T$ satisfies the FONC, SONC, and SOSC. It is a strict local minimizer.
- Actually, $\boldsymbol{x} = [0, 0]^T$ is a strict global minimizer.



Exercise 6.19

- An amphibian vehicle needs to travel from point A (on land) to point B (in water). The speeds at which the vehicle travels on land and water are v_1 and v_2, respectively.
  - Suppose that the vehicle traverses a path that minimizes the total time taken to travel from A to B. Use the FONC to show that for the optimal path above, the angles θ_1 and θ_2 satisfy Snell's law:
  $$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2}$$
  - Does the minimizer for the problem in part a satisfy the second-order sufficient condition?



Exercise 6.19

- Let x be the decision variable. Write the total travel time as $f(x)$ which is given by

$$f(x) = \frac{\sqrt{1+x^2}}{v_1} + \frac{\sqrt{1+(d-x)^2}}{v_2}$$

- Differentiating the expression

$$f'(x) = \frac{x}{v_1\sqrt{1+x^2}} - \frac{d-x}{v_2\sqrt{1+(d-x)^2}}$$

- By the first-order necessary condition, the optimal path satisfies $f'(x^*) = 0$, which corresponds to

$$\frac{x^*}{v_1\sqrt{1+(x^*)^2}} = \frac{d-x^*}{v_2\sqrt{1+(d-x^*)^2}}$$

which leads to

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2}$$

Exercise 6.19

- The second derivative of f is given by

$$f''(x) = \frac{1}{v_1(1+x^2)^{3/2}} - \frac{1}{v_2(1+(d-x)^2)^{3/2}}$$

- Hence, $f''(x) > 0$, which shows that the second order sufficient condition holds.

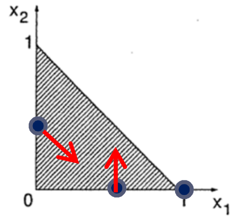
Exercise 6.20

- Suppose that you have a piece of land to sell and you have two buyers. If the first buyer receives a fraction x_1 of the piece of land, the buyer will pay you $U_1(x_1)$ dollars. Similarly, the second buyer will pay you $U_2(x_2)$ dollars for a fraction of x_2 of the land. You goal is to sell parts of your land to the two buyers so that you maximize the total dollars you receive.
  - Formulate the problem as an optimization problem of the kind

$$\begin{aligned}&\text{maximize } f(\boldsymbol{x})\\&\text{subject to } \boldsymbol{x} \in \Omega\end{aligned}$$

  - Suppose that $U_i(x_i) = a_i x_i, i = 1, 2,$ where a_1 and a_2 are given positive constants such that $a_1 > a_2$. Find all feasible points that satisfy the first-order necessary condition.
  - Among those points in the answer of part b, find all that also satisfy the second-order necessary condition.

Exercise 6.20



- We have $f(\boldsymbol{x}) = U_1(x_1) + U_2(x_2)$ and $\Omega = \{\boldsymbol{x} : x_1, x_2 \geq 0, x_1 + x_2 \leq 1\}$
- We have $\nabla f(\boldsymbol{x}) = [a_1, a_2]^T$. Because $\nabla f(\boldsymbol{x}) \neq \mathbf{0}$ for all $\boldsymbol{x}$, we conclude that no interior points satisfy the FONC. Next, consider any feasible point $\boldsymbol{x}$ for $x_2 > 0$. At such a point, the vector $\boldsymbol{d} = [1, -1]^T$ is a feasible direction. But $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = a_1 - a_2 > 0$ which means that FONC is violated (recall that the problem is to maximize f). So clearly the remaining candidates are those $\boldsymbol{x}$ for which $x_2 = 0$. Among these, if $x_1 < 1$, then $\boldsymbol{d} = [0, 1]^T$ is a feasible direction, in which case we have $\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = a_2 > 0$. This leaves the point $\boldsymbol{x} = [1, 0]^T$. At this point, any feasible direction satisfies $d_1 \leq 0$ and $d_2 \leq -d_1$. Hence, for any feasible direction, we have
$$\boldsymbol{d}^T \nabla f(\boldsymbol{x}) = d_1 a_1 + d_2 a_2 \leq d_1 a_1 + (-d_1) a_2 = d_1(a_1 - a_2) \leq 0$$
So, the only feasible point satisfies FONC is $\boldsymbol{x} = [1, 0]^T$

Exercise 6.20

- We have $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{O} \leq 0$. Hence, any point satisfies SONC (again, recall that the problem is to maximize f).