131890-S Optimization methods **Winter 2023/2024**

Lecture 2: Analytical optimization with constraints

Daniel Kaszyński *02 November 2023 r.*

2.1 Properties of gradient

We have already introduced the concept of the gradient of a function – which is a vector of partial derivatives. Gradient will be quite often used as part of a lecture, which is why we should consider its properties. As a remainder:

Definition 1: Gradient

Let $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{D}$. Through a gradient of function f, we call function $\nabla_f(x) : \mathbb{R}^n \to \mathbb{R}^n$ at point x:

$$\nabla_f(x) = \left[\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right]$$

Remember that:

Relationship between Directional Derivative and Gradient? If the gradient of a function exists $\nabla_f(\mathbf{x})$ at point $\mathbf{x}$ (which means that f is differentiable in $\mathbf{x}$)

$$\nabla_f(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right]$$

then directional derivative of a function f in direction of a vector $\mathbf{h}$ is equal to the dot product of the gradient ∇f and vector $\mathbf{h}$.

Theorem 1: Gradient is the direction of the fastest growth.

Proof. Let $|h| = 1$, i.e. let it be a normalized vector. Then the growth rate of a function f at point x in direction h is given by a directional derivative $\frac{df}{dh}(x)$. Let us determine then a direction h, which maximizes the growth rate of a function f, i.e. direction which maximizes the directional derivative:

$$\frac{df}{dh}(x) = \nabla_f(x)h = |\nabla_f(x)||h|cos(\nabla_f(x), h) = |\nabla_f(x)|cos(\nabla_f(x), h)$$

for $|\nabla_f(x)| \geq 0$ and $cos(\nabla_f(x), h) \in [-1, 1]$ the growth rate of f is the greatest when $cos(\nabla_f(x), h) = 1$, which implies that h points in the same direction that $\nabla_f(x)$ is. As a result $h = \frac{\nabla_f(x)}{|\nabla_f(x)|}$. □

Theorem 2: Gradient is orthogonal to the level set of the function.

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^* = (x_1^*, \ldots, x_n^*)$ and $\nabla_f(x^*) \neq 0$. Let $r : \mathbb{R} \to \mathbb{R}^n$ so that $r(t_0) = x^*$. Value of the function is constant for all points from a chosen level set (according to the definition) and $\forall_{t\in\mathbb{R}} f(r(t)) = c$. Then $\frac{d}{dt}(f(r(t)) = \nabla_f(r(t))\frac{dr}{dt}(t) = 0$. In particular $\nabla_f(r(t_0))\frac{dr}{dt}(t_0) = 0$.

Because $\frac{dr}{dt}(t_0)$ is a tangent space to the level set of a function f at x^*, it implies that $\nabla_f(x^*)$ is orthogonal to the level set. □

2.1.1 First Order Conditions, equality constraints

Theorem 3: Lagrange theorem, Lagrange multipliers.
Let $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$. If function f has an extremum at x related to $h(x) = 0$, where $h : \mathbb{R}^n \to \mathbb{R}^m$, $h \in C^1$, at point x, then

$$\nabla_f(x) + \lambda^T \mathbf{D}h(x) = \mathbf{0}$$

For $h : \mathbb{R}^n \to \mathbb{R}$.

$$\nabla_f(x) + \lambda\nabla_h(x) = \mathbf{0}$$

Example 1. Let $f(x) = x_1^2 + x_2^2$, and $[x_1, x_2] : h(x) = x_1^2 + 2x_2^2 - 1 = 0$. Find the extremum of $f(x)$ related to $h(x) = 1$.

From First Order Conditions (FOC) we get:

$$\begin{cases} 2x_1 + \lambda 2x_1 = 0 \Rightarrow x_1(1+\lambda) = 0 \\ 2x_2 + \lambda 4x_2 = 0 \Rightarrow x_2(1+2\lambda) = 0 \end{cases}$$

Which gives us 4 solutions:

1. $[x_1, x_2] = \left[0, \frac{1}{\sqrt{2}}\right]$, $\lambda = -\frac{1}{2}$

2. $[x_1, x_2] = \left[0, -\frac{1}{\sqrt{2}}\right]$, $\lambda = -\frac{1}{2}$

3. $[x_1, x_2] = [1, 0]$, $\lambda = -1$

4. $[x_1, x_2] = [-1, 0]$, $\lambda = -1$

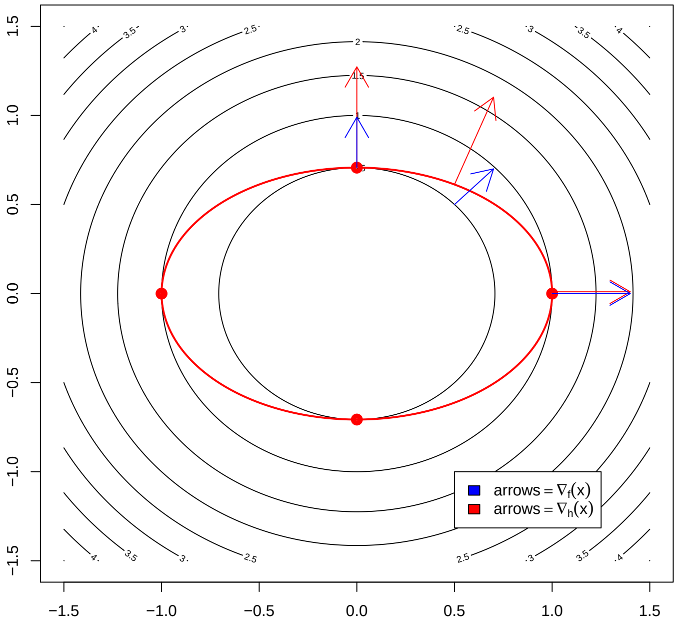

Figure 2.1: First Order Conditions equality constraints visualized

2.1.2 Second Order Conditions, equality constraints

Theorem 4: Second Order Conditions of Lagrange theorem.
Let $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$*,* $h : \mathbb{R}^n \to \mathbb{R}^m$*,* $h \in C^2$*. Let there be* x *and* λ *such that:*

1. $\nabla_f(x) + \lambda^T \mathbf{D}h(x) = 0$*, and*
2. $\forall_{z \in T(x), z \neq 0}$ *we have* $z^T H_{L(x)} z > 0$

Then by point x *we call minimum of a function* f*, related to* $h(x) = 0$*.* $T(x)$ *we call tangent space, i.e.:* $T(x) = \{z \in \mathbb{R}^n : z^T \mathbf{D}h(x) = 0\}$*. In a case when* $m = 1$*, we have* $T(x) = \{z \in \mathbb{R}^n : z^T \nabla_h(x) = 0\}$

Example 2. Lets consider 4 solutions that we have found from FOC:

1. $[x_1, x_2] = \left[0, \frac{1}{\sqrt{2}}\right]$, $\lambda = -\frac{1}{2}$

$$z : z^T \nabla_h(x) = [z_1, z_2][2x_1, 4x_2]^T = [z_1, z_2]\left[0, \frac{4}{\sqrt{2}}\right] = 0$$

$$z_1 0 + z_2 \frac{4}{\sqrt{2}} = 0 \Rightarrow \mathbf{z} = [\alpha, 0]$$

$$[\alpha,0]^T H_f([x_1,x_2])[\alpha,0] = [\alpha,0]^T \begin{bmatrix} 2+2\lambda & 0 \\ 0 & 2+4\lambda \end{bmatrix} [\alpha,0] = [\alpha,0]^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} [\alpha,0] = \alpha^2 > 0$$

2. $[x_1,x_2] = \left[0, -\frac{1}{\sqrt{2}}\right]$, $\lambda = -\frac{1}{2}$
3. $[x_1,x_2] = [1,0]$, $\lambda = -1$

$$z : z^T \nabla_h(x) = [z_1,z_2][2x_1,4x_2]^T = [z_1,z_2]\,[1,0] = 0$$

$$z_1 1 + z_2 0 = 0 \Rightarrow \mathbf{z} = [0,\alpha]$$

$$[0,\alpha]^T H_f([x_1,x_2])[0,\alpha] = [0,\alpha]^T \begin{bmatrix} 2+2\lambda & 0 \\ 0 & 4+\lambda \end{bmatrix} [0,\alpha] = [0,\alpha]^T \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} [0,\alpha] = -2\alpha^2 < 0$$

4. $[x_1,x_2] = [-1,0]$, $\lambda = -1$

2.1.3 First Order Conditions, inequality constraints

Theorem 5: KKT (Karush-Kuhn-Tucker) conditions.
Let $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$ be an objective function, and let $h : \mathbb{R}^n \to \mathbb{R}^m$, $h \in C^1$ and $g : \mathbb{R}^n \to \mathbb{R}^p$, $g \in C^1$ be the constraints. If x is an extremum then there exists a pair of $\lambda = (\lambda_1, \dots, \lambda_n)$ oraz $\mu = (\mu_1, \dots, \mu_n)$ such that:

1. ***Stationarity condition**: $\nabla_f(x) + \sum_{i=1}^m \lambda_i \nabla_{h_i(x)} + \sum_{i=1}^p \mu_i \nabla_{g_i(x)} = 0$*
2. ***Primal feasibility**: $\forall_{i=1,\dots,m} h_i(x) = 0$ and $\forall_{i=1,\dots,p} g_i(x) \le 0$*
3. ***Dual feasibility**: $\forall_{i=1,\dots,p} \mu_i \ge 0 \leftarrow$ for minimum*
4. ***Complementary slackness**: $\forall_{i=1,\dots,p} \mu_i g_i(x) = 0$*

Example 3. $f(x) = x_1^2 + x_2^2$ constrained by $[x_1,x_2] : g(x) = x_1^2 + 2x_2^2 - 1 \le 0$

From FOC we have:

$$[2x_1, 2x_2] + \mu[2x_1, 4x_2] = 0$$

$$\begin{cases} 2x_1 + \mu 2x_1 = 0 \Rightarrow x_1(1+\lambda) = 0 \\ 2x_2 + \mu 4x_2 = 0 \Rightarrow x_2(1+2\lambda) = 0 \end{cases}$$

Which gives 4 solutions:

1. $[x_1,x_2] = \left[0, \frac{1}{\sqrt{2}}\right]$, $\mu = -\frac{1}{2}$ **Dual feasibility**
2. $[x_1,x_2] = \left[0, -\frac{1}{\sqrt{2}}\right]$, $\mu = -\frac{1}{2}$ **Dual feasibility**

3. $[x_1, x_2] = [1, 0]$, $\mu = -1$ **Dual feasibility**
4. $[x_1, x_2] = [-1, 0]$, $\mu = -1$ **Dual feasibility**
5. $[x_1, x_2] = [0, 0]$, $\mu = 0$ **Stationarity condition**

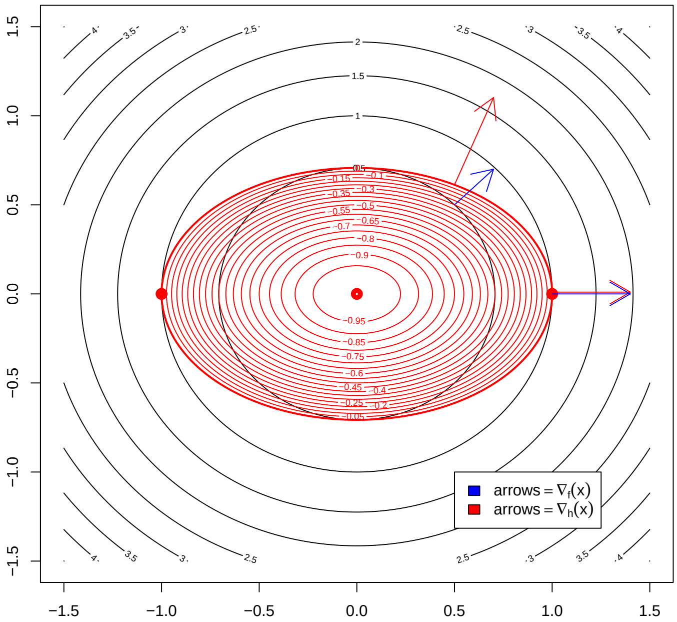

Figure 2.2: First Order Conditions inequality constraints visualized

2.1.4 Second Order Conditions, inequality constraints

Theorem 6: SOC theorem (KKT theorem).
Let $f : \mathbb{D} \subset \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $h \in C^2$, $g : \mathbb{R}^n \to \mathbb{R}^p$, $g \in C^2$. *Let there be* x, λ *and* μ, *such that:*

1. $\nabla_f(x) + \lambda^T \mathbf{D}h(x) + \mu^T \mathbf{D}g(x) = 0$, *and*

2. $\forall_{z \in T(x), z \neq 0}$ *we have* $z^T H_{L(x)} z > 0$

Then we call point x *a minimum of* f, *related to* $h(x) = 0$.

$T(x)$ *we call tangent space, i.e.:* $T(x) = \{z \in \mathbb{R}^n : z^T \mathbf{D}h(x) = 0\}$. *In a case where* $m = 1$, *we have* $T(x) = \{z \in \mathbb{R}^n : z^T \nabla_h(x) = 0\}$