KKT optimality conditions

Dimitar Dimitrov

Örebro University

May, 2011

Topics addressed in this material

- KKT (first-order) necessary conditions
  - nonlinear equality constraints
  - nonlinear inequality constraints
- second-order necessary conditions (nonlinear constraints)
- Fritz John (first-order) necessary conditions

KKT necessary conditions (linear constraints) - recap

Consider the problem

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \quad \boldsymbol{A} \in \mathbb{R}^{m_i \times n} \\ & \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}, \quad \boldsymbol{C} \in \mathbb{R}^{m_e \times n}. \end{aligned}$$

Let $\boldsymbol{x}^\star$ be a local minimizer of the above problem, where f_0 is continuously differentiable from $\mathbb{R}^n$ to $\mathbb{R}$. Then the following conditions are satisfied.

First-order necessary conditions

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x}^\star \le \boldsymbol{b}, &\quad \text{primal feasibility condition} \\ \boldsymbol{C}\boldsymbol{x}^\star = \boldsymbol{d}, &\quad \text{primal feasibility condition} \\ \boldsymbol{\lambda}^\star \ge \boldsymbol{0}, &\quad \text{dual feasibility condition} \\ (\boldsymbol{A}\boldsymbol{x}^\star - \boldsymbol{b})^T\boldsymbol{\lambda}^\star = 0, &\quad \text{complementarity condition} \\ \nabla f_0(\boldsymbol{x}^\star) + \boldsymbol{A}^T\boldsymbol{\lambda}^\star + \boldsymbol{C}^T\boldsymbol{\nu}^\star = \boldsymbol{0}, &\quad \text{stationarity condition} \end{aligned}$$

Additional assumption

If in addition, we assume that the normals to the active constraints at $\boldsymbol{x}^\star$ are linearly independent, then the Lagrange multipliers $\boldsymbol{\nu}^\star$ and $\boldsymbol{\lambda}^\star$ that satisfy the above conditions are unique [7], pp. 316.

Stationarity condition (linear constraints) - restatement

By noting that the gradients of the linear constraints $h_i(\boldsymbol{x}) = \boldsymbol{c}_i^T\boldsymbol{x} - d_i = 0$ and $f_i(\boldsymbol{x}) = \boldsymbol{a}_i^T\boldsymbol{x} - b_i \leq 0$ are given by $\nabla h_i = \boldsymbol{c}_i$ and $\nabla f_i = \boldsymbol{a}_i$, respectively,

we can restate the stationarity condition as

$$\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m_i} \lambda_i^\star \nabla f_i + \sum_{i=1}^{m_e} \nu_i^\star \nabla h_i = \boldsymbol{0},$$

or

$$\nabla f_0(\boldsymbol{x}^\star) + \sum_{i\in\mathcal{A}(\boldsymbol{x}^\star)} \lambda_i^\star \nabla f_i + \sum_{i=1}^{m_e} \nu_i^\star \nabla h_i = \boldsymbol{0},$$

where $\mathcal{A}(\boldsymbol{x}^\star)$ is a set containing the indexes of the active inequality constraints at $\boldsymbol{x}^\star$.

Using the Lagrangian function

The stationarity condition can be stated as

$$\nabla_x L(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \boldsymbol{0},$$

where $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is the Lagrangian function given by

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m_i} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{m_e} \nu_i h_i(\boldsymbol{x}).$$

KKT necessary conditions (nonlinear constraints)

Regular point

A feasible vector $\tilde{\boldsymbol{x}}$ at which the gradients of the active constraints are linearly independent is called regular [7], pp. 285.

Consider the problem

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m_i \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, m_e \end{aligned}$$

Let $\boldsymbol{x}^\star$ be a local minimizer of the above problem, where f_i and h_i are continuously differentiable from $\mathbb{R}^n$ to $\mathbb{R}$ (for all i). Then, if $\boldsymbol{x}^\star$ is regular, there exist unique Lagrange multipliers $\boldsymbol{\lambda}^\star = (\lambda_1^\star, \ldots, \lambda_{m_i}^\star)$, $\boldsymbol{\nu}^\star = (\nu_1^\star, \ldots, \nu_{m_e}^\star)$ that satisfy the following conditions [7], pp. 316.

First-order conditions

$$\begin{aligned} f_i(\boldsymbol{x}^\star) \leq 0, \quad & i = 1, \ldots, m_i, \quad \text{primal feasibility condition} \\ h_i(\boldsymbol{x}^\star) = 0, \quad & i = 1, \ldots, m_e, \quad \text{primal feasibility condition} \\ \lambda_i^\star \geq 0, \quad & i = 1, \ldots, m_i, \quad \text{dual feasibility condition} \\ \lambda_i^\star f_i(\boldsymbol{x}^\star) = 0, \quad & i = 1, \ldots, m_i, \quad \text{complementarity condition} \end{aligned}$$

$$\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m_i} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{m_e} \mu_i^\star \nabla h_i(\boldsymbol{x}^\star) = \mathbf{0}, \quad \text{stationarity condition.}$$

KKT necessary conditions (nonlinear constraints) - summary

The problem

$$\begin{aligned} \underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, m_i \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, \dots, m_e \end{aligned}$$

First-order conditions

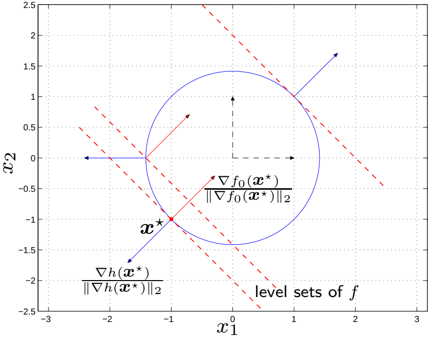
$$\begin{aligned} f_i(\boldsymbol{x}^\star) \leq 0, \quad & i = 1, \dots, m_i, \quad && \text{primal feasibility condition} \\ h_i(\boldsymbol{x}^\star) = 0, \quad & i = 1, \dots, m_e, \quad && \text{primal feasibility condition} \\ \lambda_i^\star \geq 0, \quad & i = 1, \dots, m_i, \quad && \text{dual feasibility condition} \\ \lambda_i^\star f_i(\boldsymbol{x}^\star) = 0, \quad & i = 1, \dots, m_i, \quad && \text{complementarity condition} \end{aligned}$$

$$\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m_i} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{m_e} \mu_i^\star \nabla h_i(\boldsymbol{x}^\star) = \mathbf{0}, \quad \text{stationarity condition.}$$

KKT (necessary) conditions for $\boldsymbol{x}^\star$ to be a local minimizer for the problem

- $\boldsymbol{x}^\star$ is regular
- there exist unique Lagrange multipliers $\boldsymbol{\lambda}^\star = (\lambda_1^\star, \dots, \lambda_{m_i}^\star)$, $\boldsymbol{\nu}^\star = (\nu_1^\star, \dots, \nu_{m_e}^\star)$ that satisfy the first-order conditions

Example (nonlinear constraints)



Consider the problem ([3], pp. 308, [7], pp. 284)

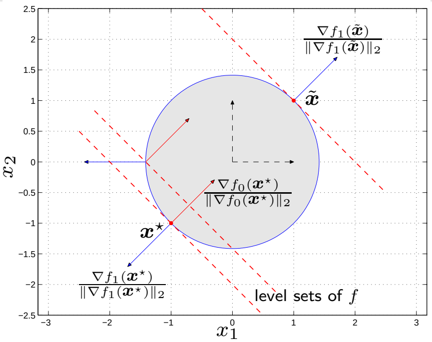
$$\underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} \quad f_0(\boldsymbol{x}) = x_1 + x_2$$

$$\text{subject to} \quad h(\boldsymbol{x}) = x_1^2 + x_2^2 - 2 = 0.$$

This is a problem with a linear objective function $f(\boldsymbol{x})$ and one nonlinear equality constraint $h(\boldsymbol{x}) = 0$. At the solution $\boldsymbol{x}^\star$, the gradient of the constraint $\nabla h(\boldsymbol{x}^\star)$ is orthogonal to the level set of the function at $\boldsymbol{x}^*$, and hence $\nabla h(\boldsymbol{x}^\star)$ and $\nabla f_0(\boldsymbol{x}^\star)$ are parallel $i.e.$, there is a scalar $\nu^\star$ such that

$$\nabla f_0(\boldsymbol{x}^\star) + \nu^\star \nabla h(\boldsymbol{x}^\star) = \mathbf{0}.$$

Clearly, in this example $\boldsymbol{x}^\star$ is regular (because $\nabla h(\boldsymbol{x}^\star) \neq \mathbf{0}$).



Consider the problem ([3], pp. 310)

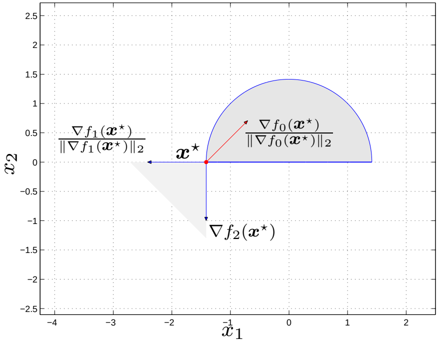
$$\underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} \;\; f_0(\boldsymbol{x}) = x_1 + x_2$$

$$\text{subject to} \;\; f_1(\boldsymbol{x}) = x_1^2 + x_2^2 - 2 \leq 0.$$

This is a problem with a linear objective function $f(\boldsymbol{x})$ and one nonlinear inequality constraint $f_1(\boldsymbol{x}) \leq 0$. At the solution $\boldsymbol{x}^\star$, the gradient of the constraint $\nabla f_1(\boldsymbol{x}^\star)$ is orthogonal to the level set of the function at $\boldsymbol{x}^*$, and the following equality holds

$$\nabla f_0(\boldsymbol{x}^\star) + \lambda^\star \nabla f_1(\boldsymbol{x}^\star) = \mathbf{0},$$

for $\lambda^\star = \frac{1}{2} \geq 0$. Note that at the point $\tilde{\boldsymbol{x}} = (1, 1)$, $\nabla f_0(\tilde{\boldsymbol{x}}) + \lambda \nabla f_1(\tilde{\boldsymbol{x}}) = \mathbf{0}$ holds as well, however $\lambda = -\frac{1}{2} \leq 0$.



Consider the problem ([3], pp. 313)

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} \quad & f_0(\boldsymbol{x}) = x_1 + x_2 \\ \text{subject to} \quad & f_1(\boldsymbol{x}) = x_1^2 + x_2^2 - 2 \le 0, \\ & f_2(\boldsymbol{x}) = -x_2 \le 0. \end{aligned}$$

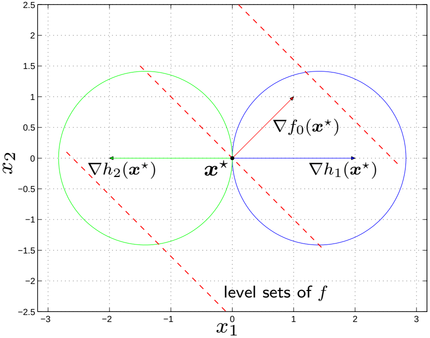
At the solution $\boldsymbol{x}^\star = (-\sqrt{2}, 0)$, $-\nabla f_0(\boldsymbol{x}^\star)$ belongs to the normal cone to the feasible set at point $\boldsymbol{x}^\star$, hence, there is $\boldsymbol{\lambda}^\star \ge \mathbf{0}$ that satisfies

$$\nabla f_0(\boldsymbol{x}^\star) + \lambda_1^\star \nabla f_1(\boldsymbol{x}^\star) + \lambda_2^\star \nabla f_2(\boldsymbol{x}^\star) = \mathbf{0}.$$

We saw that when dealing only with linear constraints, the regularity of $\boldsymbol{x}^\star$ affects only the uniqueness of the Lagrange multipliers. In the general case (when nonlinear constraints are present), however, it can have additional implications (as we show with the following three examples).

Consider the problem ([2], pp. 78)

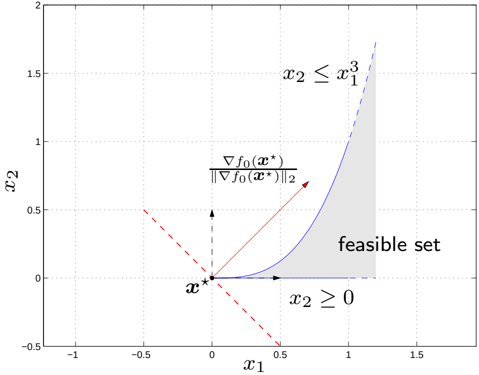
$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} \quad & f_0(\boldsymbol{x}) = x_1 + x_2 \\ \text{subject to} \quad & h_1(\boldsymbol{x}) = (x_1+1)^2 + x_2^2 - 2 = 0. \\ & h_2(\boldsymbol{x}) = (x_1-1)^2 + x_2^2 - 2 = 0. \end{aligned}$$



This is a problem with a linear objective function $f_0(\boldsymbol{x})$ and two nonlinear equality constraints $h_1(\boldsymbol{x}) = 0$, $h_2(\boldsymbol{x}) = 0$. At the solution $\boldsymbol{x}^\star$ (which is in fact the only feasible point), no linear combination of the gradients of the two constraints is equal to $\nabla f_0(\boldsymbol{x}^\star)$, $i.e$, there is no $\boldsymbol{\nu}^\star$ such that

$$\nabla f_0(\boldsymbol{x}^\star) + \nu_1^\star \nabla h_1(\boldsymbol{x}^\star) + \nu_2^\star \nabla h_2(\boldsymbol{x}^\star) = \mathbf{0}.$$

Hence, without the assumption that $\boldsymbol{x}^\star$ is a *regular* point, the first-order conditions are clearly <u>not</u> *necessary* for $\boldsymbol{x}^\star$ to be a local minimizer.

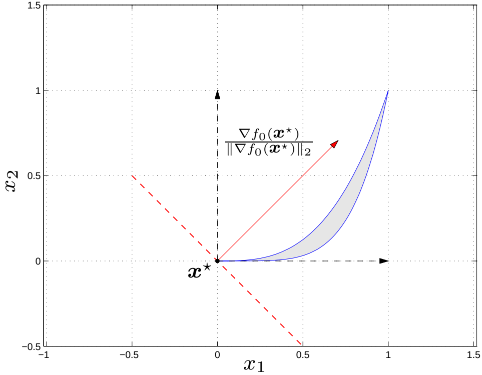


Consider the problem ([5], pp. 203)

$$\begin{aligned} \underset{\boldsymbol{x} \in \mathbb{R}^2}{\text{minimize}} \quad & f_0(\boldsymbol{x}) = x_1 + x_2 \\ \text{subject to} \quad & f_1(\boldsymbol{x}) = x_2 - x_1^3 \leq 0, \\ & f_2(\boldsymbol{x}) = -x_2 \leq 0. \end{aligned}$$

This is a problem with a linear objective function $f_0(\boldsymbol{x})$, one nonlinear inequality constraint $f_1(\boldsymbol{x}) \leq 0$ and one linear inequality constraint $f_2(\boldsymbol{x}) \leq 0$. There are infinitely many feasible points, however, at $\boldsymbol{x}^\star$ (where both inequality constraints are active), there is no $\boldsymbol{\lambda}^\star$ satisfying

$$\nabla f_0(\boldsymbol{x}^\star) + \lambda_1^\star \nabla f_1(\boldsymbol{x}^\star) + \lambda_2^\star \nabla f_2(\boldsymbol{x}^\star) = \mathbf{0}.$$

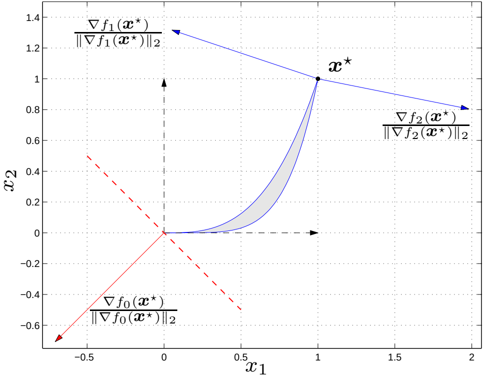


Consider the problem ([6], pp. 116)

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} \quad & f_0(\boldsymbol{x}) = x_1 + x_2 \\ \text{subject to} \quad & f_1(\boldsymbol{x}) = x_2 - x_1^3 \leq 0, \\ & f_2(\boldsymbol{x}) = x_1^5 - x_2 \leq 0, \\ & f_3(\boldsymbol{x}) = -x_2 \leq 0. \end{aligned}$$

At the solution $\boldsymbol{x}^\star$, all inequality constraints are active and their gradients are co-linear. There is no $\boldsymbol{\lambda}^\star$ satisfying

$$\nabla f_0(\boldsymbol{x}^\star) + \lambda_1^\star \nabla f_1(\boldsymbol{x}^\star) + \lambda_2^\star \nabla f_2(\boldsymbol{x}^\star) + \lambda_3^\star \nabla f_3(\boldsymbol{x}^\star) = \mathbf{0}.$$



$$
\begin{aligned}
\underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{maximize}} \quad & f_0(\boldsymbol{x}) = x_1 + x_2 \\
\text{subject to} \quad & f_1(\boldsymbol{x}) = x_2 - x_1^3 \leq 0, \\
& f_2(\boldsymbol{x}) = x_1^5 - x_2 \leq 0, \\
& f_3(\boldsymbol{x}) = -x_2 \leq 0.
\end{aligned}
$$

Only the first two inequality constraints are active at the solution $\boldsymbol{x}^* = (1, 1)$, which satisfies the KKT necessary conditions with $\lambda_1^\star = 3$, $\lambda_2^\star = 2$ and $\lambda_3^\star = 0$. This can be verified by solving the equation

$$
\begin{bmatrix} \nabla f_1(x_1) \\ \nabla f_2(x_1) \end{bmatrix}^T \begin{bmatrix} \lambda_1^\star \\ \lambda_2^\star \end{bmatrix} = -\nabla f(\boldsymbol{x}^\star) \Rightarrow \begin{bmatrix} -3 & 1 \\ 5 & -1 \end{bmatrix}^T \begin{bmatrix} \lambda_1^\star \\ \lambda_2^\star \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.
$$

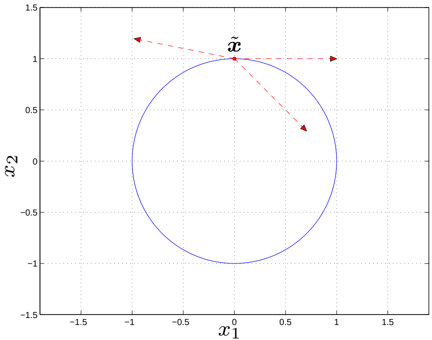
In order the answer the above question, we need to revisit our definition of **feasible directions**.

Feasible direction [6], pp. 88

Suppose that we are at a point $\tilde{\boldsymbol{x}} \in \mathcal{S} \subseteq \mathbb{R}^n$. $\Delta \boldsymbol{x} \in \mathbb{R}^n$ defines a *feasible direction* at $\tilde{\boldsymbol{x}}$, if a "small" step in the direction $\Delta \boldsymbol{x}$ does not lead outside of the set $\mathcal{S}$. In other words,

$$\exists \delta > 0 \text{ such that } \tilde{\boldsymbol{x}} + \alpha \Delta \boldsymbol{x} \in \mathcal{S},$$

for all $\alpha \in [0, \delta]$.



Wait a minute ... feasible directions?

In the general case, when dealing with nonlinear constraints, there may be no feasible directions, as defined above. The figure depicts an equality constraint

$$h(\boldsymbol{x}) = x_1^2 + x_2^2 - 1 = 0.$$

Even an infinitesimally small step in any direction from a feasible point $\tilde{\boldsymbol{x}}$ would lead us outside of the set $\mathcal{S}$ ($i.e.$, the boundary of the unit circle).

Feasible curves - nonlinear constraints

It is convenient to define vector valued functions $h(\boldsymbol{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{m_e}$, $f(\boldsymbol{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{m_i}$

$$h(\boldsymbol{x}) = \begin{bmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_{m_e}(\boldsymbol{x}) \end{bmatrix}, \quad f(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_{m_i}(\boldsymbol{x}) \end{bmatrix}.$$

The set of points $\boldsymbol{x}$ such that $h(\boldsymbol{x}) = \mathbf{0}$ is called a surface.

In order to retain feasibility with respect to a set of nonlinear equality constraints, we have to move on the surface defined by them. We call a curve on which the function $h(\boldsymbol{x})$ remain identically equal to zero, a **feasible curve** (with respect to the equality constraints).

Directed curve (also known as an arc).

We define an arc passing through a point $\tilde{\boldsymbol{x}}$ as a set of points

$$\{\alpha(\theta) : \theta \in [\underline{\theta}, \overline{\theta}]\},$$

parametrized by a single variable θ ranging from $\underline{\theta}$ to $\overline{\theta}$. We assume that $\alpha(0) = \tilde{\boldsymbol{x}}$. A **feasible arc** should satisfy for all θ

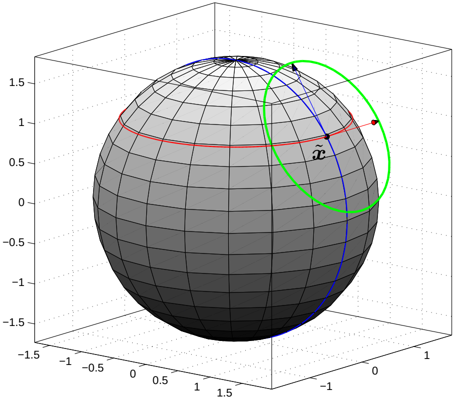
$$h(\alpha(\theta)) = \mathbf{0}, \quad f(\alpha(\theta)) \leq \mathbf{0}.$$

We will consider only curves that are continuously differentiable, $i.e.$, $\frac{d(\alpha(\theta))}{d\theta}$ exists.

Consider the constraint

$$h(\boldsymbol{x}) = x_1^2 + x_2^2 + x_3^2 - 3 = 0.$$

There are infinitely many arcs passing through the feasible point $\tilde{\boldsymbol{x}} = (\sqrt{2}, 0, 1)$. Some of them are feasible arcs. Two such feasible arcs are depicted on the figure.



$$\alpha_1(\theta_1) = \begin{bmatrix} \sqrt{2}\cos(\theta_1) \\ \sqrt{2}\sin(\theta_1) \\ 1 \end{bmatrix},$$

$$\alpha_2(\theta_2) = \begin{bmatrix} \sqrt{3}\cos(\theta_2 + \gamma) \\ 0 \\ \sqrt{3}\sin(\theta_2 + \gamma) \end{bmatrix},$$

where $\gamma = \frac{1}{\sqrt{3}}$ is an offest so that $\alpha_2(0) = \tilde{\boldsymbol{x}}$, and $\theta_1, \theta_2 \in [-\pi, \pi]$.

One can readily verify that $h(\alpha_1(\theta_1)) = h(\alpha_2(\theta_2)) = 0$. The tangents to both curves at $\tilde{\boldsymbol{x}}$ are depicted as well. They are given by

$$\dot{\alpha}_1(0) = \left.\frac{d\alpha_1(\theta_1)}{d\theta_1}\right|_{\theta_1=0} = \begin{bmatrix} 0 \\ \sqrt{2} \\ 0 \end{bmatrix}, \quad \dot{\alpha}_2(0) = \left.\frac{d\alpha_2(\theta_2)}{d\theta_2}\right|_{\theta_2=0} = \begin{bmatrix} -\sqrt{3}\sin(\gamma) \\ 0 \\ \sqrt{3} \end{bmatrix}.$$

If $\boldsymbol{x}^\star$ is a local minimizer of f_0, then it is a local minimizer of f_0 along any feasible arc passing through $\boldsymbol{x}^\star$ [8], pp. 515.

Suppose that $\alpha(\theta)$ is any such arc, with $\alpha(0) = \boldsymbol{x}^\star$. Then, if $\theta = 0$ is a local minimizer of the one-dimensional function $f_0(\alpha(\theta))$, the derivative of $f_0(\alpha(\theta))$ with respect to θ must vanish at $\theta = 0$. Using the chain rule leads to

$$\left.\frac{df_0(\alpha(\theta))}{d\theta}\right|_{\theta=0} = \left.\nabla f_0(\alpha(\theta))^T \dot{\alpha}(\theta)\right|_{\theta=0} = \nabla f_0(\boldsymbol{x}^\star)^T \dot{\alpha}(0) = 0.$$

Let us define the set of all tangents to feasible arcs through $\boldsymbol{x}^\star$

$$\mathcal{T}(\boldsymbol{x}^\star) := \{\boldsymbol{p} : \boldsymbol{p} = \dot{\alpha}(0), \text{for some feasible arc } \alpha(\theta) \text{ through } \boldsymbol{x}^\star\}.$$

With the help of $\mathcal{T}(\boldsymbol{x}^\star)$ we can state the following optimality condition

$\boldsymbol{x}^\star$ is a local minimizer of f_0 if

$$\nabla f_0(\boldsymbol{x}^\star)^T \boldsymbol{p} = 0, \quad \text{for all } \boldsymbol{p} \in \mathcal{T}(\boldsymbol{x}^\star).$$

The above condition is not practical, since (in general) it is not easy to represent the set of all feasible arcs explicitly.

The tangent cone

$\mathcal{T}(\boldsymbol{x}^\star)$ is called the *tangent cone*

If we assume (for convenience) that $\mathbf{0} \in \mathcal{T}(\boldsymbol{x}^\star)$, then the set $\mathcal{T}(\boldsymbol{x}^\star)$ is a cone. This can be seen by noting that if $\boldsymbol{p} \in \mathcal{T}(\boldsymbol{x}^\star)$, so does $\mu\boldsymbol{p}$, for some nonnegative μ. We require for μ to be nonnegative, because our arc has *direction*.

By noting that "a tangent is a limiting direction of a feasible sequence" [3], pp. 316, we can define the tangent cone in an alternative way

Tangent cone [7], pp. 343.

Given a subset $\mathcal{X}$ of $\mathbb{R}^n$ and a vector $\boldsymbol{x} \in \mathcal{X}$, a vector $\boldsymbol{p}$ is said to be a tangent of $\mathcal{X}$ at $\boldsymbol{x}$ if either $\boldsymbol{p} = \mathbf{0}$ or there exists a sequence $\{\boldsymbol{x}^{(k)}\} \subset \mathcal{X}$ such that $\boldsymbol{x}^{(k)} \neq \boldsymbol{x}$ for all k and

$$\boldsymbol{x}^{(k)} \to \boldsymbol{x}, \qquad \frac{\boldsymbol{x}^{(k)} - \boldsymbol{x}}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|} \to \frac{\boldsymbol{p}}{\|\boldsymbol{p}\|}.$$

The set of all tangents of $\mathcal{X}$ at $\boldsymbol{x}$ is called the **tangent cone** of $\mathcal{X}$ at $\boldsymbol{x}$.

In words, the above definition states that a nonzero vector $\boldsymbol{p}$ is tangent to $\mathcal{X}$ at $\boldsymbol{x}$ if it is possible to approach $\boldsymbol{x}$ with a feasible sequence $\{\boldsymbol{x}^{(k)}\}$, such that the normalized **direction sequence** $\boldsymbol{x}^{(k)} - \boldsymbol{x}$ converges to the normalized direction $\boldsymbol{p}$.

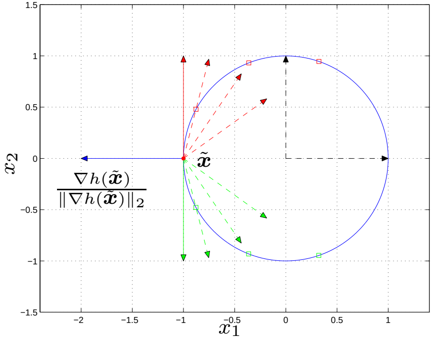
Consider again the equality constraint

$$h(\boldsymbol{x}) = x_1^2 + x_2^2 - 1 = 0.$$

The figure shows two feasible sequences approaching $\tilde{\boldsymbol{x}}$ (one depicted in red and the other one in green). The green and red arrows represent

$$\frac{\boldsymbol{x}^{(k)} - \boldsymbol{x}}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|},$$

for some choices of feasible points $\boldsymbol{x}^{(k)}$ (depicted with green and red squares).



Any vector of the form $\boldsymbol{p} = (0, \delta)$, where $\delta \in \mathbb{R}$, is tangent to a feasible sequence at $\tilde{\boldsymbol{x}}$.

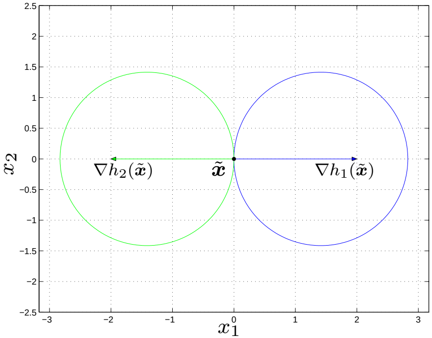
Observation

In this particular example, the tangent cone $\mathcal{T}(\tilde{\boldsymbol{x}})$ is equivalent to the set of vectors $\boldsymbol{v}$ that satisfy

$$\nabla h(\tilde{\boldsymbol{x}})^T \boldsymbol{v} = 0.$$

Or in other words, $\mathcal{T}(\tilde{\boldsymbol{x}}) = \mathcal{N}(\nabla h(\tilde{\boldsymbol{x}})^T)$.

Consider again the example where only one feasible point $\tilde{\boldsymbol{x}} = (0, 0)$ exists. In this case, the tangent cone $\mathcal{T}(\tilde{\boldsymbol{x}}) = \mathbf{0}$, since there are no feasible sequences that converge to $\tilde{\boldsymbol{x}}$.



Only the vector $\boldsymbol{p} = (0, 0)$, is "tangent" to a "feasible sequence" at $\tilde{\boldsymbol{x}}$.

Observation

In this particular example, the tangent cone $\mathcal{T}(\tilde{\boldsymbol{x}})$ is **not** equivalent to the set of vectors $\boldsymbol{v}$ that satisfy

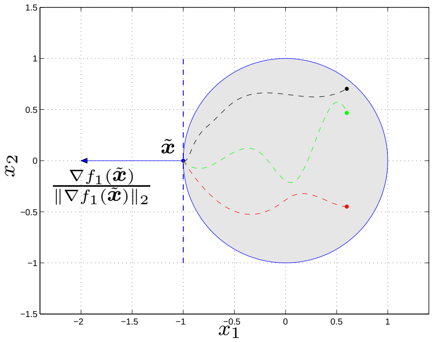
$$\begin{bmatrix} \nabla h_1(\tilde{\boldsymbol{x}})^T \\ \nabla h_2(\tilde{\boldsymbol{x}})^T \end{bmatrix} \boldsymbol{v} = \mathbf{0}.$$

Consider again the feasible set defined by the constraint

$$f_1(\boldsymbol{x}) = x_1^2 + x_2^2 - 1 \leq 0.$$

Three feasible sequences approaching $\tilde{\boldsymbol{x}}$ are depicted. Any feasible arc must satisfy

$$f_1(\alpha(\theta)) \leq 0, \quad \text{for all } \theta.$$



All vectors of the form $\boldsymbol{p} = (\delta_1, \delta_2)$, where $\delta_1 \in \mathbb{R}_+$, are tangent to a feasible sequence at $\tilde{\boldsymbol{x}}$.

Observation

In this particular example, the tangent cone $\mathcal{T}(\tilde{\boldsymbol{x}})$ is equivalent to the set of vectors $\boldsymbol{v}$ that satisfy

$$\nabla f_1(\tilde{\boldsymbol{x}})^T \boldsymbol{v} \leq \mathbf{0}.$$

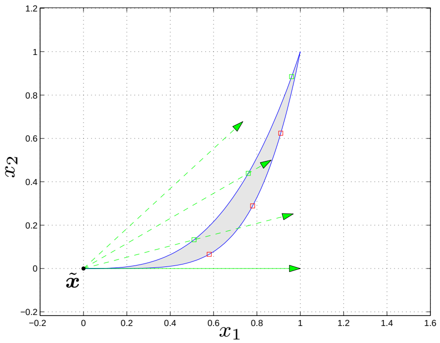
Example (tangent cone, inequality constraints)

Consider again the feasible set in the figure. Two feasible sequences approaching $\tilde{\boldsymbol{x}}$ are depicted. The green arrows represent

$$\frac{\boldsymbol{x}^{(k)} - \boldsymbol{x}}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|},$$

for some choices of feasible points $\boldsymbol{x}^{(k)}$ (depicted with green and red squares). The red arrows are omitted. Any feasible arc must satisfy

$$f_1(\alpha(\theta)) \leq 0, \quad \text{for all } \theta.$$



Only vectors of the form $\boldsymbol{p} = (\delta, 0)$, where $\delta \in \mathbb{R}_+$, are tangent to a feasible sequence at $\tilde{\boldsymbol{x}}$.

Observation

In this particular example, the tangent cone $\mathcal{T}(\tilde{\boldsymbol{x}})$ is **not** equivalent to the set of vectors $\boldsymbol{v}$ that satisfy

$$\begin{bmatrix} \nabla f_1(\tilde{\boldsymbol{x}})^T \\ \nabla f_2(\tilde{\boldsymbol{x}})^T \\ \nabla f_3(\tilde{\boldsymbol{x}})^T \end{bmatrix} \boldsymbol{v} \leq \boldsymbol{0}.$$

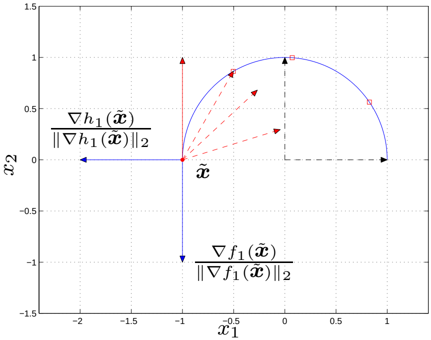
Consider the feasible set defined by the constraints

$$h_1(\boldsymbol{x}) = x_1^2 + x_2^2 - 1 = 0,$$
$$f_1(\boldsymbol{x}) = -x_2 \leq 0.$$

The figure shows one feasible sequence approaching $\tilde{\boldsymbol{x}}$ (depicted in red). The red arrows represent

$$\frac{\boldsymbol{x}^{(k)} - \boldsymbol{x}}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|},$$

for some choices of feasible points $\boldsymbol{x}^{(k)}$ (depicted with red squares).



Any vector of the form $\boldsymbol{p} = (0, \delta)$, where $\delta \in \mathbb{R}_+$, is tangent to a feasible sequence at $\tilde{\boldsymbol{x}}$.

Observation

In this particular example, the tangent cone $\mathcal{T}(\tilde{\boldsymbol{x}})$ is equivalent to the set of vectors $\boldsymbol{v}$ that satisfy both

$$\nabla h_1(\tilde{\boldsymbol{x}})^T \boldsymbol{v} = 0,$$
$$\nabla f_1(\tilde{\boldsymbol{x}})^T \boldsymbol{v} \leq 0.$$

The tangent cone is not very amendable to manipulation [5], pp. 202, and it is convenient to consider a way to approximate it. Our observations so far suggest that except for some "special cases" it seems that a linear approximation is reasonable.

Define a set of linearized constraints at point $\tilde{\boldsymbol{x}}$

Let the i^{th} row of matrix $\boldsymbol{C}$ contain the gradient of the i^{th} equality constraint, and the i^{th} row of matrix $\hat{\boldsymbol{A}}$ contain the gradient of the i^{th} active inequality constraint at $\tilde{\boldsymbol{x}}$ $i.e.$,

$$\boldsymbol{C}(\tilde{\boldsymbol{x}}) = \begin{bmatrix} \nabla h_1(\tilde{\boldsymbol{x}})^T \\ \vdots \\ \nabla h_{m_e}(\tilde{\boldsymbol{x}})^T \end{bmatrix} \in \mathbb{R}^{m_e \times n}, \quad \hat{\boldsymbol{A}}(\tilde{\boldsymbol{x}}) = \begin{bmatrix} \nabla f_{i \in \mathcal{A}(\tilde{\boldsymbol{x}})}(\tilde{\boldsymbol{x}})^T \end{bmatrix} \in \mathbb{R}^{|\mathcal{A}(\tilde{\boldsymbol{x}})| \times n},$$

where $|\mathcal{A}(\tilde{\boldsymbol{x}})|$ denotes the cardinality of the set $\mathcal{A}(\tilde{\boldsymbol{x}})$ ($i.e.$, the number of active inequality constraints at $\tilde{\boldsymbol{x}}$). Define a set $\mathcal{F}(\tilde{\boldsymbol{x}})$ of linearized constraints at point $\tilde{\boldsymbol{x}}$ as

$$\mathcal{F}(\tilde{\boldsymbol{x}}) = \{\boldsymbol{p} \in \mathbb{R}^n : \boldsymbol{C}(\tilde{\boldsymbol{x}})\boldsymbol{p} = \boldsymbol{0}, \hat{\boldsymbol{A}}(\tilde{\boldsymbol{x}})\boldsymbol{p} \leq \boldsymbol{0}\}, \quad \text{[3], pp. 316.}$$

Note that the set $\mathcal{F}(\tilde{\boldsymbol{x}})$ is a cone (why?).

$\mathcal{T}(\tilde{\boldsymbol{x}}) \subseteq \mathcal{F}(\tilde{\boldsymbol{x}})$

In all our examples we observed that the tangent cone at $\tilde{\boldsymbol{x}}$ is a subset of $\mathcal{F}(\tilde{\boldsymbol{x}})$. This is indeed the case in general [5], pp. 202.

When does $\mathcal{T}(\tilde{x}) = \mathcal{F}(\tilde{x})$ hold?

$\mathcal{T}(\tilde{x}) = \mathcal{F}(\tilde{x})$ holds if [5], pp. 203

- All active constraints at $\tilde{\boldsymbol{x}}$ are linear.
- All active constraints at $\tilde{\boldsymbol{x}}$ have linearly independent gradients ($i.e.$, $\tilde{\boldsymbol{x}}$ is a regular point).

Essentially, the regularity assumptions guarantees that the linear approximation $\mathcal{F}(\tilde{x})$ captures the essential geometric features of the feasible set near $\tilde{\boldsymbol{x}}$ [3], pp. 315.

We can think of $\mathcal{F}(\tilde{x})$ as a set (in fact a subspace) of **first order feasible variations**. Let us consider for example the i^{th} equality constraint $h_1(\boldsymbol{x}) = 0$. The Taylor-series expansion about a feasible point $\tilde{\boldsymbol{x}}$ along the direction $\boldsymbol{p}$ gives

$$h_1(\tilde{\boldsymbol{x}} + \boldsymbol{p}) \approx h_1(\tilde{\boldsymbol{x}}) + \nabla h_1(\tilde{\boldsymbol{x}})^T \boldsymbol{p} = \nabla h_1(\tilde{\boldsymbol{x}})^T \boldsymbol{p}.$$

The constraint $h_1(\tilde{\boldsymbol{x}} + \boldsymbol{p}) = 0$ retains feasibility, **to first-order**, when it satisfies $\nabla h_1(\tilde{\boldsymbol{x}})^T \boldsymbol{p} = 0$. In case of an inequality constraint $f_1(\boldsymbol{x}) \leq 0$ we have

$$f_1(\tilde{\boldsymbol{x}} + \boldsymbol{p}) \approx f_1(\tilde{\boldsymbol{x}}) + \nabla f_1(\tilde{\boldsymbol{x}})^T \boldsymbol{p} = \nabla f_1(\tilde{\boldsymbol{x}})^T \boldsymbol{p}.$$

Hence, using a similar argument, we conclude that $\nabla f_1(\tilde{\boldsymbol{x}})^T \boldsymbol{p} \leq 0$ is desirable.

Regularity assumption (again) - only nonlinear equality constraints

Worth noting ... (1)

Suppose that there are only nonlinear equality constraints. Again, let the i^{th} row of matrix $\boldsymbol{C}$ contain the gradient of the i^{th} equality constraint, *i.e.*,

$$\boldsymbol{C}(\boldsymbol{x}^\star) = \begin{bmatrix} \nabla h_1(\boldsymbol{x}^\star)^T \\ \vdots \\ \nabla h_{m_e}(\boldsymbol{x}^\star)^T \end{bmatrix} \in \mathbb{R}^{m_e \times n}.$$

Then, if $\boldsymbol{x}^\star$ is regular, *i.e.*, $\text{rank}(\boldsymbol{C}(\boldsymbol{x}^\star)) = m_e$, $\mathcal{T}(\boldsymbol{x}^\star) = \mathcal{N}(\boldsymbol{C}(\boldsymbol{x}^\star))$.

Of course, in this case the set $\mathcal{F}(\tilde{\boldsymbol{x}})$ is simply given by

$$\mathcal{F}(\tilde{\boldsymbol{x}}) = \{\boldsymbol{p} \in \mathbb{R}^n : \boldsymbol{C}(\boldsymbol{x}^\star)\boldsymbol{p} = \boldsymbol{0}\}.$$

Worth noting ... (2)

For a nice note regarding the "**true set of feasible variations**" $\mathcal{T}(\tilde{\boldsymbol{x}})$ and **first-order feasible variations** $\mathcal{F}(\tilde{\boldsymbol{x}})$ see [7], pp. 287.

[3], pp. 319

"Constraint qualifications are conditions under which the linearized feasible set $\mathcal{F}(\tilde{\boldsymbol{x}})$ is similar to the tangent cone $\mathcal{T}(\tilde{\boldsymbol{x}})$. In fact, most constraint qualifications ensure that these two sets are identical. As mentioned earlier, these conditions ensure that the $\mathcal{F}(\tilde{\boldsymbol{x}})$, which is constructed by linearizing the algebraic description of the feasible set at $\tilde{\boldsymbol{x}}$, captures the essential geometric features of the feasible set in the vicinity of $\tilde{\boldsymbol{x}}$, as represented by $\mathcal{T}(\tilde{\boldsymbol{x}})$."

The condition that the gradients of the active constraints at $\tilde{\boldsymbol{x}}$ are linearly independent, is known as the Linear Independence Constraint Qualification (LICQ). There is a variety of other constraint qualifications that vary from abstract (and difficult to check) to more specific (and easyly verifiable), but somewhat restrictive in many situations [6], pp. 112. LICQ is a good example of the latter type of constraint qualifications (it could pose strong assumptions in many practical situations [6], pp. 132).

We will encounter the so called **Slater's constraint qualification** when we deal with convex optimization problems.

"It is important to note that the definition of tangent cone does not rely on the algebraic specification of the feasible set, only on its geometry. The linearized feasible direction set does, however, depend on the definition of the constraint functions."
[3], pp. 316

Example

Let

$$h(x_1, x_2) = x_1 = 0.$$

This equality constraint yields the x_2 axis, and every point on that axis is regular because the gradient at each point is equal to $\nabla h = (1, 0)$. If instead we define

$$h(x_1, x_2) = x_1^2 = 0,$$

again the feasible set is the x_2 axis but now no point on in is regular, because $\nabla h = (2x_1, 0) = (0, 0)$, since $x_1 = 0$. [9], pp. 325

Second-order necessary conditions [8], pp. 519

Here, we assume that f_i, h_i (for all i) are twice continuously differentiable.

Recall that

if $\boldsymbol{x}^\star$ is a local minimizer of f_0, then it is a local minimizer of f_0 along any feasible arc passing through $\boldsymbol{x}^\star$.

Suppose that $\alpha(\theta)$ is any such arc, with $\alpha(0) = \boldsymbol{x}^\star$. Then, if $\theta = 0$ is a local minimizer of the one-dimensional function $f_0(\alpha(\theta))$, the second derivative of $f_0(\alpha(\theta))$ with respect to θ must be nonnegative at $\theta = 0$. Using the chain rule leads to

$$\frac{d^2 f_0(\alpha(\theta))}{d\theta^2} = \frac{\nabla f_0(\alpha(\theta))^T \dot{\alpha}(\theta)}{d\theta} = \dot{\alpha}(\theta)^T \nabla^2 f_0(\alpha(\theta))\dot{\alpha}(\theta) + \nabla f_0(\alpha(\theta))^T \ddot{\alpha}(\theta) \geq 0.$$

Hence, at $\theta = 0$ (by using $\alpha(0) = \boldsymbol{x}^\star$ and $\boldsymbol{p} = \dot{\alpha}(0)$) we have

$$\frac{d^2 f_0(\boldsymbol{x}^\star)}{d\theta^2} = \boldsymbol{p}^T \nabla^2 f_0(\boldsymbol{x}^\star)\boldsymbol{p} + \nabla f_0(\boldsymbol{x}^\star)^T \ddot{\alpha}(0) \geq 0. \tag{1}$$

"The second derivative along an arc depends not only on the Hessian of the objective, but also on the curvature of the constraints (that is, on the term $\ddot{\alpha}(0)$)"

Getting rid of $\ddot{\alpha}(0)$

For simplicity of presentation, we assume that there are only equality constraints (however, the results can be extended to inequality constraints in a straightforward way)

Note that

since $h(\theta)$ is constant along any feasible arc, its second derivative with respect to θ must vanish for all θ (and in particular at $\theta = 0$).

$$\frac{d^2 h_i(\alpha(\theta))}{d\theta^2} = \frac{\nabla h_i(\alpha(\theta))^T \dot{\alpha}(\theta)}{d\theta} = \dot{\alpha}(\theta)^T \nabla^2 h_i(\alpha(\theta))\dot{\alpha}(\theta) + \nabla h_i(\alpha(\theta))^T \ddot{\alpha}(\theta) = 0.$$

Hence, at $\theta = 0$ we have

$$\frac{d^2 h_i(\boldsymbol{x}^\star)}{d\theta^2} = \boldsymbol{p}^T \nabla^2 h_i(\boldsymbol{x}^\star)\boldsymbol{p} + \nabla h_i(\boldsymbol{x}^\star)^T \ddot{\alpha}(0) = 0.$$

Next, we multiply by $\nu_i^\star$ and sum over all i to obtain

$$\sum_{i=1}^{m_e} \nu_i \frac{d^2 h_i(\boldsymbol{x}^\star)}{d\theta^2} = \boldsymbol{p}^T \sum_{i=1}^{m_e} \nu_i \nabla^2 h_i(\boldsymbol{x}^\star)\boldsymbol{p} + \underbrace{\sum_{i=1}^{m_e} \nu_i \nabla h_i(\boldsymbol{x}^\star)^T}_{-\nabla f_0(\boldsymbol{x}^\star)^T} \ddot{\alpha}(0) = 0. \qquad (2)$$

Since, from the first-order conditions we have that $\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m_e} \nu_i \nabla h_i(\boldsymbol{x}^\star) = 0.$

Summing (1) and (2) leads to

$$\boldsymbol{p}^T \left(\nabla^2 f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m_e} \nu_i \nabla^2 h_i(\boldsymbol{x}^\star) \right) \boldsymbol{p} \geq 0.$$

for all tangent vectors $\boldsymbol{p} \in \mathcal{T}(\boldsymbol{x}^\star)$. The matrix in the brackets is the Hessian of the Lagrangian function with respect to $\boldsymbol{x}$ at the point $(\boldsymbol{x}^\star, \boldsymbol{\nu}^\star)$, *i.e.*,

$$\nabla^2_{xx} L(\boldsymbol{x}^\star, \boldsymbol{\nu}^\star) = \nabla^2 f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m_e} \nu_i \nabla^2 h_i(\boldsymbol{x}^\star),$$

where the Lagrangian function is given by

$$L(\boldsymbol{x}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m_e} \nu_i h_i(\boldsymbol{x}).$$

Note that since $\boldsymbol{p} \in \mathcal{T}(\boldsymbol{x}^\star)$, it follows that $\boldsymbol{p} \in \mathcal{N}(\boldsymbol{C}(\boldsymbol{x}^\star))$, hence we can express $\boldsymbol{p}$ as

$$\boldsymbol{p} = \boldsymbol{Z}(\boldsymbol{x}^\star)\boldsymbol{p}_z,$$

for some $\boldsymbol{p}_z$, where the columns of $\boldsymbol{Z}(\boldsymbol{x}^\star)$ span $\mathcal{N}(\boldsymbol{C}(\boldsymbol{x}^\star))$.

Projected Hessian of the Lagrangian

Two equivalent conditions:

$$\boldsymbol{p}^T \nabla_{xx}^2 L(\boldsymbol{x}^\star, \boldsymbol{\nu}^\star)\boldsymbol{p} \geq 0.$$

for all tangent vectors $\boldsymbol{p} \in \mathcal{T}(\boldsymbol{x}^\star)$.

$$\boldsymbol{p}_z^T \boldsymbol{Z}(\boldsymbol{x}^\star)^T \nabla_{xx}^2 L(\boldsymbol{x}^\star, \boldsymbol{\nu}^\star)\boldsymbol{Z}(\boldsymbol{x}^\star)\boldsymbol{p}_z \geq 0.$$

for all $\boldsymbol{p}_z \in \mathbb{R}^z$, where z is the dimension of the null space of $\boldsymbol{C}(\boldsymbol{x}^\star)$.

Second-order necessary condition

The matrix $\boldsymbol{Z}(\boldsymbol{x}^\star)^T \nabla_{xx}^2 L(\boldsymbol{x}^\star, \boldsymbol{\nu}^\star)\boldsymbol{Z}(\boldsymbol{x}^\star)$ is called the *projected Hessian of the Lagrangian*, and it should be positive semidefinite at a local minimizer $\boldsymbol{x}^\star$.

Note that the second-order necessary conditions for a nonlinear constrained problem depend on a "special combination" of the Hessian of the objective function and Hessian matrices of the constraints.

It is interesting to see that we can formulate necessary conditions for optimality without the regularity assumption. We illustrate this on the following nonlinear inequality constrained problem

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0, \quad i = 1,\dots,m_i. \end{aligned}$$

If $\boldsymbol{x}^\star$ is a local minimizer for the above problem, then there exist multipliers $\lambda_i^\star$, $i = 0,\dots,m_i$ (not all equal to zero), such that

$$\begin{aligned} \sum_{i=0}^{m_i} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) &= \mathbf{0}, \\ \lambda_i^\star f_i(\boldsymbol{x}^\star) = 0, &\quad i = 1,\dots,m_i, \\ \lambda_i^\star \geq 0, &\quad i = 0,\dots,m_i. \end{aligned}$$

For a proof (using the Farkas' lemma) see [6], pp. 122.

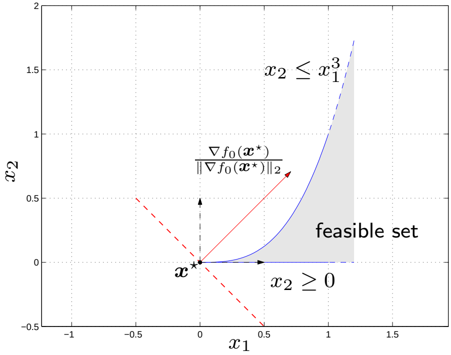
Essentially, a multiplier $\lambda_0^\star$ is assigned to the gradient of the objective function, so (loosely speaking) in cases when $\boldsymbol{x}^\star$ is not regular, $\lambda_0^\star = 0$ can be chosen. Of course setting $\lambda_0^\star = 0$ means that the objective function plays no role in the optimality conditions - this is a rather unexpected (and generally speaking, unwanted situation) [6]. pp. 124.

Consider again the problem ([5], pp. 203)

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} \quad & f_0(\boldsymbol{x}) = x_1 + x_2 \\ \text{subject to} \quad & f_1(\boldsymbol{x}) = x_2 - x_1^3 \le 0, \\ & f_2(\boldsymbol{x}) = -x_2 \le 0. \end{aligned}$$

Clearly, the solution $\boldsymbol{x}^\star$ is not regular, and there is no $\boldsymbol{\lambda}^\star = (\lambda_1^\star, \lambda_2^\star)$ satisfying

$$\nabla f_0(\boldsymbol{x}^\star) + \lambda_1^\star \nabla f_1(\boldsymbol{x}^\star) + \lambda_2^\star \nabla f_2(\boldsymbol{x}^\star) = \mathbf{0}.$$



The Fritz John necessary conditions

$$\lambda_0^\star \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\nabla f_0(\boldsymbol{x}^\star)} + \lambda_1^\star \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\nabla f_1(\boldsymbol{x}^\star)} + \lambda_2^\star \underbrace{\begin{bmatrix} 0 \\ -1 \end{bmatrix}}_{\nabla f_2(\boldsymbol{x}^\star)} = 0, \quad \lambda_0, \lambda_1, \lambda_2 \ge 0, \quad (\text{not all } \lambda_i^\star \text{ equal to zero}),$$

are satisfied for $\lambda_0^\star = 0$, and $\lambda_1^\star = \delta$, $\lambda_2^\star = \delta$, for any $\delta > 0$. Note that the complementarity condition is satisfied because all constraints are active at $\boldsymbol{x}^\star$.

[1] S. Boyd, and L. Vandenberghe, "Convex Optimization," Cambridge, 2004.

[2] P. E. Gill, W. Murray, and M. H. Wright, "Practical Optimization," Emerald, 2007.

[3] J. Nocedal, and S. J. Wright, "Numerical Optimization," Springer.

[4] D. Bertsimas, and J. N. Tsitsiklis, "Introduction to Linear Optimization," Athena Scientific, 1997.

[5] R. Fletcher, "Practical Methods of Optimization," Wiley, (2^{rd} edition), 1987.

[6] N. Andréasson, A. Evgrafov, and M. Patriksson, "An Introduction to Continuous Optimization: Foundations and Fundamental Algorithms," 2005.

[7] D. P. Bertsekas, "Nonlinear Programming," Athena Scientific, (3^{rd} print) 2008.

[8] I. Griva, S. G. Nash, A. Sofer, "Linear and Nonlinear Optimization," SIAM, 2009.

[9] D. G. Luenberger, Y. Ye, "Linear and Nonlinear Programming," Springer, 2010.