Introduction to Local and Global Optimization for NLP

Marco Trubian

Dipartimento di Scienze dell'Informazione (DSI)
Università degli Studi di Milano

Constrained Optimization

We want to minimize functions subject to constraints on the variables

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) & \leq 0 & j = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array}$$

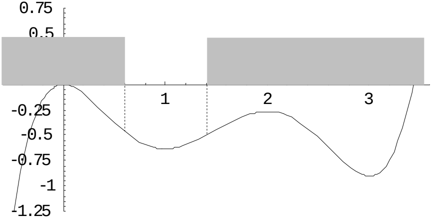
with $\boldsymbol{x} \in \mathbb{R}^n$.

Example I

In presence of constraints a difficult problem can become easy:

$$f(x) = \frac{1}{5}x^5 - \frac{3}{2}x^4 + \frac{11}{3}x^3 - 3x^2.$$

is not convex in $\mathbb{R}$, while it is convex, e.g. in the interval $X = [0.6, 1.4]$. Starting from $x_0 \in X$ any scalar optimization technique would find the global optimum.

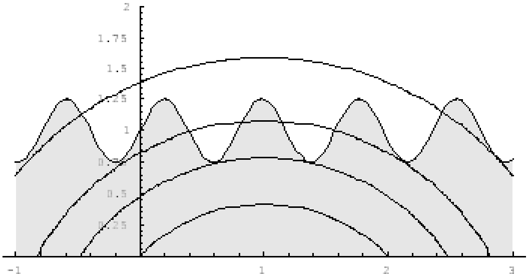


Example II

Let us consider the following problem

$$\begin{aligned} \min \quad & f(x,y) = (x-1)^2 + (y+1)^2 \\ & g_1(x,y) = 1 + \tfrac{1}{4}\sin(8x) - y \leq 0; \\ & g_2(x,y) = -y \leq 0. \end{aligned}$$

- The objective function f is convex and allows just one stationary point $(1,-1)$
- $(1,-1)$ optimum of the unconstrained problem
- the constrained problem has an infinite number of local minima



Optimality conditions: equality constraints

Let us consider the following problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ h_j(\boldsymbol{x}) & = 0 \quad j = 1, \ldots, h < n \end{array}$$

and its *Lagrangean function*

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{h} \lambda_j h_j(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda^T \boldsymbol{h}(\boldsymbol{x}).$$

First order optimality conditions

Optimality conditions can be given by means of the Lagrangean function

Theorem We are given a function $f(\boldsymbol{x})$ and h equality constraints $h_j(\boldsymbol{x}) = 0$, with $j = 1, \ldots, h$, with $f(\boldsymbol{x})$ and h_j of class C^1. *Under the hypothesis that the vectors* $\nabla h_j(\boldsymbol{x}^*)$ *are linearly indipendent*, if $\boldsymbol{x}^*$ is a local minimum of $f(\boldsymbol{x})$ which satisfies the equality constraints, then there exists $\boldsymbol{\lambda}^*$ s.t. $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a stationary point of the Lagrangean function $L(\boldsymbol{x}, \boldsymbol{\lambda})$:

$$\frac{\partial L}{\partial x_i} = \frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} + \sum_{j=1}^{h} \lambda_j^* \frac{\partial h_j(\boldsymbol{x}^*)}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n \tag{1}$$

$$\frac{\partial L}{\partial \lambda_j} = h_j(\boldsymbol{x}^*) = 0, \quad j = 1, 2, \ldots, h \tag{2}$$

First order optimality conditions

The conditions

$$\frac{\partial L}{\partial x_i} = \frac{\partial f(\boldsymbol{x})}{\partial x_i} + \sum_{j=1}^{h} \lambda_j \frac{\partial h_j(\boldsymbol{x})}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n \qquad (3)$$

$$\frac{\partial L}{\partial \lambda_j} = h_j(\boldsymbol{x}^*) = 0, \quad j = 1, 2, \ldots, h \qquad (4)$$

are a system of $n + h$ equations in $n + h$, $\boldsymbol{x}, \lambda$, unknowns.

The first n conditions can be written as $\nabla f(\boldsymbol{x}^*) + J(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^* = \mathbf{0}$, or

$$-\nabla f(\boldsymbol{x}^*) = \sum_{j=1}^{h} \lambda_j^* \nabla h_j(\boldsymbol{x}^*)$$

i.e., in a stationary point $\boldsymbol{x}^*$ the antigradient of f is given by a linear combination of the gradient vectors of the equality constraints

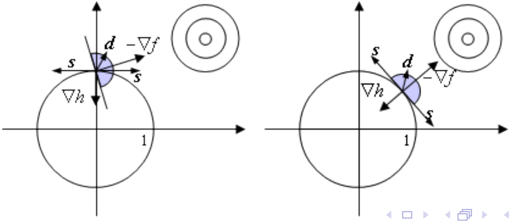
Example

Example Given the problem

$$\min \quad f(x,y) = (x-2)^2 + (y-2)^2$$
$$h_1(x,y) = 1 - x^2 - y^2 = 0.$$

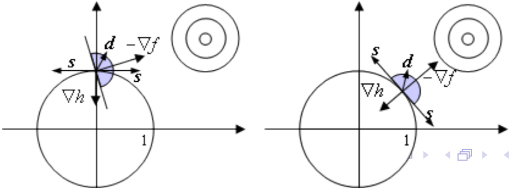
In the optimum point $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$

- the antigradient of f, $-\nabla f(x,y) = -(2(x-2), 2(y-2))^T$, is the vector $(4-\sqrt{2}, 4-\sqrt{2})^T$
- the gradient of h_1, $\nabla h(x,y) = (-2x, -2y)^T$, is $-(\sqrt{2}, \sqrt{2})$
- in the figure on the right the to vectors are collinear



Geometric interpretation

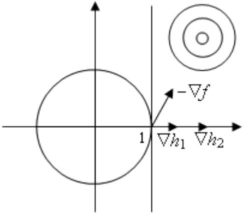
On the left: point $(0, 1)$
Hyperplane $F = \{\boldsymbol{s} \in \mathbb{R}^n \text{ s.t. } \nabla h(x, y)^T \boldsymbol{s} = 0\}$
F = first order approximation of $h(x, y)$
Subspace $D = \{\boldsymbol{d} \in \mathbb{R}^n \text{ s.t. } \nabla f(x, y)^T \boldsymbol{d} < 0\}$
D = all descent directions (shadowed halfcircle)
Point not optimal: there are descent directions which belong to F
(along them, at least for a infinitesimal distance, we improve f while satisfying the equality constraint)
On the right: point $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$
It is optimal since no descent direction belongs to a F
Here the vectors $-\nabla f(x, y)$ and $\nabla h(x, y)$ are collinear.



On constraints qualification

The First order optimality conditions are valid

> [...] *Under the hypothesis that the vectors $\nabla h_j(\boldsymbol{x}^*)$ are linearly indipendent* [...]

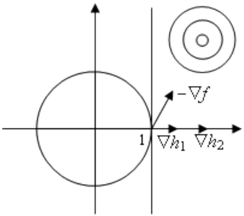


Example Given the problem

$$\min \quad f(x,y) = (x-2)^2 + (y-2)^2$$
$$h_1(x,y) = x^2 + y^2 - 1 = 0;$$
$$h_2(x,y) = x - 1 = 0.$$

The optimum is the only feasible point $(1,0)$.

On constraints qualification



In $(1,0)$ the vectors $\nabla h_1(x,y)$ and $\nabla h_2(x,y)$ are $(2,0)^T$ and $(1,0)^T$: they are linearly dependent.
In $(1,0)$ the antigradient $-\nabla f(x,y)$ is $(2,4)^T$
The system of equations has no solutions for λ_1 and λ_2

$$\begin{pmatrix} 2 \\ 4 \end{pmatrix} = \lambda_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Constraints Qualification

- When considering unconstrained problems, all local minima satisfy necessary optimality conditions, and theoretically all local minima can be found among stationary points.
- When considering constrained problems, not necessarily all local minima can be found among those which satisfy analitical conditions,

$$-\nabla f(\boldsymbol{x}^*) = \sum_{j=1}^{h} \lambda_j^* \nabla h_j(\boldsymbol{x}^*)$$

- but only those which satisfy the so called *Constraints Qualification*

Constraints Qualification

Constraints Qualification A point $\boldsymbol{x}^*$ satisfies constraint qualifications if there exists a vector $\boldsymbol{h}$ s.t. $\nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{h} < 0$, for all indeces j s.t. $g_j(\boldsymbol{x}^*) = 0$, $\nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{h} = 0$ with $j = 1, 2, \ldots, h$ and vectors $\nabla h_j(\boldsymbol{x}^*)$ with $j = 1, 2, \ldots, h$ are linearly indipendent

- Constraints Qualification must hold both for equality constraints and for the inequality constraints which are active (i.e. satisfied as equality) in $\boldsymbol{x}^*$

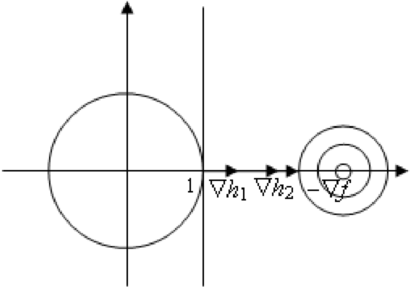
Constraints Qualification

Constraints Qualification are satisfied if:

- the set of equality constraints and active inequality constraints gradients are linearly independent in $\boldsymbol{x}^*$; (as in the theorem)
- if all constraints are linear
- if all constraints are convex and the feasible reagion has at least one internal point

Definition A point $\boldsymbol{x}^*$ which satisfies Constraints Qualification is called regular

On constraints qualification



$$\begin{aligned} \min \quad & f(x,y) = (x-3)^2 + y^2 \\ & h_1(x,y) = x^2 + y^2 - 1 = 0; \\ & h_2(x,y) = x - 1 = 0. \end{aligned}$$

in this problem the system can have solution even if the vectors $\nabla h_j(\boldsymbol{x}^*)$ are linearly dipendent. This occur since $-\nabla f(\boldsymbol{x}^*)$ can be generated by a linear combination of a subset of the vectors $\nabla h_j(\boldsymbol{x}^*)$.

First order sufficient conditions for convex problems

First order sufficient conditions

Theorem We are given a function $f(\boldsymbol{x})$ and h equality constraints $h_j(\boldsymbol{x}) = 0$, with $j = 1, \ldots, h$, with $f(\boldsymbol{x})$ and h_j convex functions of class C^1. *Under the hypothesis that the Jacobian matrix $J(\boldsymbol{x}^*)$ has full rank h,* if there exists $\boldsymbol{\lambda}^*$ s.t. $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a stationary point of the Lagrangean function $L(\boldsymbol{x}, \boldsymbol{\lambda})$ then $\boldsymbol{x}^*$ is a local minimum of $f(\boldsymbol{x})$.

Quadratic model and linear equality constraints

Let us consider the special case of a quadratica model, with Q p.d., under linear equality constraints

$$\begin{array}{ll} \min & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T\mathbf{x} \\ t.c. & A\mathbf{x} = \mathbf{d} \end{array}$$

A is a full rank $(h \times n)$ matrix, with $h < n$.
With linear constraints, Constraints Qualification are satisfied.

Quadratic model and linear equality constraints

The lagrangean function is

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x} + \boldsymbol{\lambda}^T (\boldsymbol{d} - A\boldsymbol{x})$$

First order optimality condition for $\boldsymbol{x}^*$ to be a minimum is that there exisits $\boldsymbol{\lambda}^*$ s.t.:

$$\begin{aligned} \nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= Q\boldsymbol{x}^* - \boldsymbol{b} - A^T \boldsymbol{\lambda}^* = \boldsymbol{0} \\ \nabla_\lambda L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= A\boldsymbol{x}^* - \boldsymbol{d} = \boldsymbol{0} \end{aligned}$$

which can be written as

$$\begin{bmatrix} Q & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{d} \end{bmatrix}$$

with solution

$$\begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} Q & -A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{d} \end{bmatrix}$$

From inequality to equality constraints

Let us consider the general problem

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) & \leq 0 & j = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array}$$

First technique: from $g_i(\boldsymbol{x}) \leq 0$ to $g_i(\boldsymbol{x}) + \theta_i^2 = 0$.
(Why do we square θ ?)

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) + \theta_j^2 & = 0 & j = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array}$$

with lagrangean model:

$$L(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j (g_j(\boldsymbol{x}) + \theta_j^2) + \sum_{j=1}^{h} \mu_j h_j(\boldsymbol{x})$$

From inequality to equality constraints

First order necessary optimality conditions for $\boldsymbol{x}$ are

$$
\begin{aligned}
\frac{\partial L}{\partial x_i} &= \frac{\partial f(\boldsymbol{x})}{\partial x_i} + \sum_{j=1}^{k} \lambda_j \frac{\partial g_j(\boldsymbol{x})}{\partial x_i} + \sum_{j=1}^{h} \mu_j \frac{\partial h_j(\boldsymbol{x})}{\partial x_i} = 0, \quad i = 1, 2, \dots, n \\
\frac{\partial L}{\partial \theta_j} &= 2\lambda_j \theta_j = 0, \quad j = 1, 2, \dots, k \\
\frac{\partial L}{\partial \lambda_j} &= g_j(\boldsymbol{x}) + \theta_j^2 = 0, \quad j = 1, 2, \dots, k \\
\frac{\partial L}{\partial \mu_j} &= h_j(\boldsymbol{x}) = 0, \quad j = 1, 2, \dots, h
\end{aligned}
$$

A $(n + 2k + h) \times (n + 2k + h)$ system

The k relations $2\lambda_j\theta_j = 0$, with $j = 1, 2, \dots, k$, are complementary slackness conditions: $\lambda_j = 0$ when the constraint $g_j(\boldsymbol{x}) \leq 0$ is satified as a strict inequality, and $g_j(\boldsymbol{x}) = 0$ when $\lambda_j \neq 0$

The general case: KKT conditions

Let us consider the general problem

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) & \leq 0 & j = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array} \tag{5}$$

and its lagrangean model:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j g_j(\boldsymbol{x}) + \sum_{j=1}^{h} \mu_j h_j(\boldsymbol{x})$$

The general case: KKT conditions

Theorem We are given a general problem, where the functions $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ and $h_j(\boldsymbol{x})$ are in C^1. If $\boldsymbol{x}^*$ is a local minimum and in $\boldsymbol{x}^*$ constraints qualification hold for equality and active constraints, then there are Lagrange multiplier vector $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$, s.t. the following conditions are satisfied,

$$
\begin{array}{lll}
\frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} + & \sum_{j=1}^{k} \lambda_j^* \frac{\partial g_j(\boldsymbol{x}^*)}{\partial x_i} + \sum_{j=1}^{h} \mu_j^* \frac{\partial h_j(\boldsymbol{x}^*)}{\partial x_i} = 0, & i = 1, \ldots, n \\
g_j(\boldsymbol{x}^*) & \leqslant 0, & j = 1, \ldots, k \\
\lambda_j^* g_j(\boldsymbol{x}^*) & = 0, & j = 1, \ldots, k \\
h_j(\boldsymbol{x}^*) & = 0, & j = 1, \ldots, h \\
\lambda_j^* & \geq 0, & j = 1, \ldots, k
\end{array}
$$

These conditions are often known as the Karush-Kuhn-Tucker conditions, or KKT conditions for short.

Constrained Optimization

The k conditions $\lambda_j^* g_j(\boldsymbol{x}^*) = 0$, with $j = 1, 2, \ldots, k$, are complementarity conditions; they imply that either constraint i is active or $\lambda_j^* = 0$ or possibly both. In particular, the Lagrange multipliers corresponding to inactive inequality constraints are zero.

If $I \subseteq \{1, 2, \ldots, k\}$ denotes the subset of indices $1, 2, \ldots, k$, of active inequality constraints, we can rewrite the first n conditions as

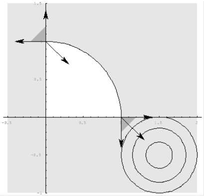
$$-\nabla f(\boldsymbol{x}^*) = \sum_{j \in I} \lambda_j^* \nabla g_j(\boldsymbol{x}^*) + \sum_{j=1}^{h} \mu_j^* \nabla h_j(\boldsymbol{x}^*) \tag{6}$$

i.e.

> in a stationary point $\boldsymbol{x}^*$ the antigradient of f is given by a nonnegative linear combination of the gradient vectors of the active inequality constraints and of a linear combination of the gradient vectors of the equality constraints

Example

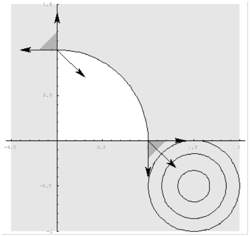
$$
\begin{aligned}
\min \quad & f(x,y) = (x-1.5)^2 + (y+0.5)^2 \\
& g_1(x,y) = -x \leq 0; \\
& g_2(x,y) = -y \leq 0; \\
& g_3(x,y) = x^2 + y^2 - 1 \leq 0.
\end{aligned}
$$



In the optimum $\boldsymbol{x}^* = (1,0)$

- the constraints g_2 e g_3 are active, and the vectors $\nabla g_2(x,y)$ and $\nabla g_3(x,y)$, $(0,-1)^T$ and $(2,0)^T$, are linearly independent
- $-\nabla f(x,y)$, is $(1,-1)^T$ and
- $\boldsymbol{\lambda}^*$ is $(0,1,1/2)^T$.
- $-\nabla f(x,y)$ belongs to the cone given by the nonnegative linear combination of the gradients of active constraints in $\boldsymbol{x}^*$

Example

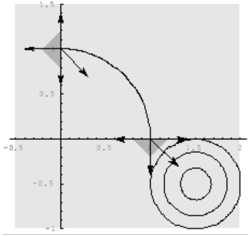


In the generic point $\boldsymbol{x} = (0, 1)$

- the constraints g_1 e g_3 are active, and the vectors $\nabla g_1(x, y)$ and $\nabla g_3(x, y)$, $(-1, 0)^T$ and $(2, 0)^T$, are linearly independent
- $-\nabla f(x, y)$, is $(3, -3)^T$ and
- $\boldsymbol{\lambda}$ is $(-3, 0, -3/2)^T$ and violates the non negativity conditions
- $-\nabla f(x, y)$ lies outside the cone given by the nonnegative linear combination of the gradients of active constraints in $\boldsymbol{x}$

Example

$$
\begin{aligned}
\min \quad & f(x,y) = (x-1.5)^2 + (y+0.5)^2 \\
& g_1(x,y) = -x \le 0; \\
& g_2(x,y) = -y \le 0; \\
& h_1(x,y) = x^2 + y^2 - 1 = 0;
\end{aligned}
$$



In the optimum $\boldsymbol{x}^* = (1,0)$

- $\boldsymbol{\lambda}^*$ is $(0,1)$ and $\mu_1 = 1/2$.
- $-\nabla f(x,y)$ belongs to the cone given by the nonnegative linear combination of the gradients of active inequality constraints and by the linear combination of the gradients of equality constraints

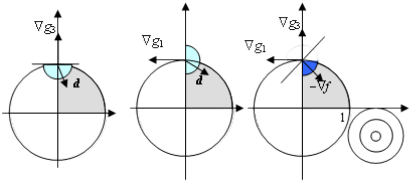
Here even a negative value for μ_1 would have been acceptable

Feasible direction

Definition Given a feasible point $\boldsymbol{x}$ we call feasible direction set the set

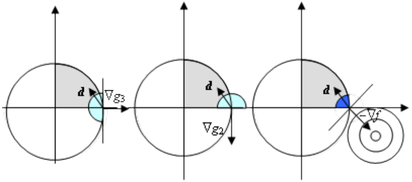
$$F(\boldsymbol{x}) = \{\boldsymbol{d} \mid \nabla h_j(\boldsymbol{x})^T \boldsymbol{d} = 0,\ j = 1, \ldots, h;\ \nabla g_j(\boldsymbol{x})^T \boldsymbol{d} \leq 0,\ j \in I\}.$$

Geometric interpretation



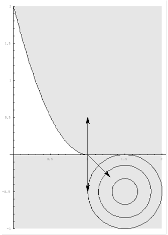
Example The feasible region X is the grey quarter of circle. Let us consider the not optimal point $(0,1)$. Since the inequality constraints are in the form of $\leq$, their gradients point outside X. Following for a small distance the directions $\boldsymbol{d}$ s.t. $\nabla g_j(\boldsymbol{x})^T \boldsymbol{d} \leq 0$ we stay within X. Such direction are the blue cone. $(0,1)$ is not an optimum since the blue cone contains descent directions

Geometric interpretation



On the contrary, $(1, 0)$ is an optimum since the blue cone (the intersection of the halfspaces of the feasible directions of the active constraints in the point) does not contain descent directions

A non regular point



$$\begin{aligned} \min \quad & f(x,y) = (x-1.5)^2 + (y+0.5)^2 \\ & g_1(x,y) = -2(x-1)^3 + y \leq 0; \\ & g_2(x,y) = -y \leq 0. \end{aligned}$$

In the optimal point $\boldsymbol{x}^* = (1,0)$

- the constraints g_1 e g_2 are active, and the vectors $\nabla g_1(x,y)$ and $\nabla g_2(x,y)$, $(0,1)^T$ and $(0,-1)^T$, are linearly dependent
- $F(\boldsymbol{x}^*) = \{(d,0)^T \mid d \in \mathbb{R}\}$
- $-\nabla f(x,y)$, is $(1,-1)^T$ and
- no $\boldsymbol{\lambda}^*$ can exist

The convex case

Theorem We are given a general problem, where the functions $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ and $h_j(\boldsymbol{x})$ are in C^1. If $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ and $h_j(\boldsymbol{x})$ are convex functions then KKT conditions are sufficient conditions.

Second order optimality conditions

Definition Given a feasible point $\boldsymbol{x}^*$ and moltiplier vectors $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ which satisfy KKT conditions, we call critical cone the set

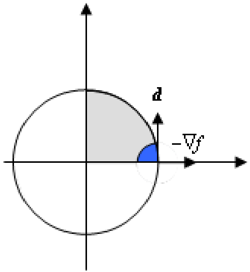
$$C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \{\boldsymbol{d} \in F(\boldsymbol{x}^*)\, |\nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} = 0,\ j \in E;\ \nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{d} = 0,\ j \in I,\ \text{with } \lambda_j^* > 0\}.$$

From KKT conditions we obtain

$$-\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} = \sum_{j \in I} \lambda_j^* \nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{d} + \sum_{j=1}^{h} \mu_j^* \nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} = 0 \quad \forall \boldsymbol{d} \in C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*).$$

The directions belonging to the critical cone are orthogonal to $\nabla f(\boldsymbol{x})$

Second order optimality conditions



In the problem

$$
\begin{aligned}
\min \quad & f(x, y) = (x - 1.5)^2 + y^2 \\
& g_1(x, y) = -x \leq 0; \\
& g_2(x, y) = -y \leq 0; \\
& g_3(x, y) = x^2 + y^2 - 1 \leq 0;
\end{aligned}
$$

The critical cone is $C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \{(1, d)^T \mid d \geq 0\}$.

Second order optimality conditions

Theorem We are given a general problem, where the functions $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ and $h_j(\boldsymbol{x})$ are in C^2. If $\boldsymbol{x}^*$ is a local minimum and in $\boldsymbol{x}^*$ constraints qualification hold for equality and active constraints, and the Lagrange multiplier vector $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ satisfy the KKT conditions, then

$$\boldsymbol{d}^T \nabla^2_{x,x} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^* \boldsymbol{\mu}^*)\boldsymbol{d} \geq 0 \;\; \forall \boldsymbol{d} \in C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*).$$

where the hessian matrix of the lagrangean function is

$$\nabla^2_{x,x} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^* \boldsymbol{\mu}^*) = H(\boldsymbol{x}^*) + \sum_{j \in I} \lambda_j^* \nabla^2 g_j(\boldsymbol{x}^*) + \sum_{j=1}^{h} \mu_j^* \nabla^2 h_j(\boldsymbol{x}^*).$$

We are requiring the semidefinite potiveness of the hessian matrix of the lagrangean function in the critical cone

Second order sufficient optimality conditions

If the hessian matrix of the lagrangean function is positive definite in the critical cone then the KKT conditions become sufficient

Theorem We are given a general problem, where the functions $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ and $h_j(\boldsymbol{x})$ are in C^2. If $\boldsymbol{x}^*$ is feasible point and the Lagrange multiplier vector $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ satisfy the KKT conditions, and the following relation holds

$$\boldsymbol{d}^T \nabla^2_{\boldsymbol{x},\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^* \boldsymbol{\mu}^*) \boldsymbol{d} \; > \; 0 \;\; \forall \boldsymbol{d} \in C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

then $\boldsymbol{x}^*$ is a strict local minimum

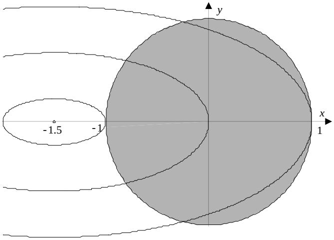
Please observe that here we no more require constraints qualification.

Second order sufficient optimality conditions

Example

$$\min \quad f(x,y) = 2(x+1.5)^2 + 10y^2$$
$$g_1(x,y) = 1 - x^2 - y^2 \leq 0;$$

A global minimum at $(-1.5, 0)^T$ where g_1 is not active and $\lambda^* = 0$, and a strict local minimum $\tilde{\boldsymbol{x}} = (1,0)$ where $g_1(1,0) = 0$.



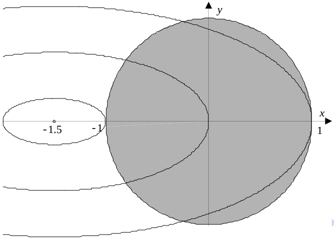
Second order sufficient optimality conditions

In the strict local minimum $\tilde{\boldsymbol{x}} = (1, 0)$ the KKT conditions hold

$$\begin{pmatrix} 4(x+1.5) \\ 20y \end{pmatrix} - \lambda_1 \begin{pmatrix} 2x \\ 2y \end{pmatrix} = \begin{pmatrix} 10 \\ 0 \end{pmatrix} - \lambda_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

with $\lambda_1^* = 5$.
The hessian matrix of the lagrangean function is

$$\nabla^2_{xx} L(\tilde{\boldsymbol{x}}, \lambda_1^*) = \begin{pmatrix} 4 & 0 \\ 0 & 20 \end{pmatrix} - \lambda_1^* \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 - 2\lambda_1^* & 0 \\ 0 & 20 - 2\lambda_1^* \end{pmatrix} = \begin{pmatrix} -6 & 0 \\ 0 & 10 \end{pmatrix}$$



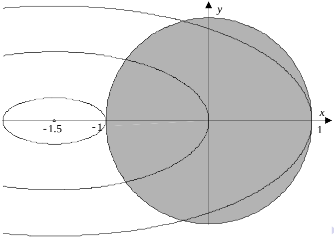
Second order sufficient optimality conditions

In $\tilde{\mathbf{x}}$, $\nabla g_1(\tilde{\mathbf{x}}) = (2,0)^T$ and the critical cone is $C(\tilde{\mathbf{x}}, \lambda_1^*) = \{(0,d)^T \mid d \in \mathbb{R}\}$.

Hence we obtain,

$$\boldsymbol{d}^T \nabla^2_{\boldsymbol{x},\boldsymbol{x}} L(\tilde{\boldsymbol{x}}, \lambda^*)\boldsymbol{d} = \begin{pmatrix} 0 \\ d \end{pmatrix}^T \begin{pmatrix} -6 & 0 \\ 0 & 10 \end{pmatrix} = \begin{pmatrix} 0 \\ d \end{pmatrix} = 10d^2 > 0.$$

In $\tilde{\boldsymbol{x}}$, second order sufficient optimality conditions hold and then $\tilde{\boldsymbol{x}}$ is a strict local minimum.



Quadratic model with linear inequality constraints

$$\begin{array}{ll} \min & q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} \\ t.c. & A\boldsymbol{x} \geq \boldsymbol{d} \end{array}$$

$\boldsymbol{x}^* = \boldsymbol{x}_0 = -Q^{-1}b$ when $A\boldsymbol{x}^* \geq \boldsymbol{d}$.
Quite relevant problem: it is iterativelly solved as a subproblem by some optimization algorithms.
If we know the set of active constraints then we reduce to the case of a quadratic model with equality constraints: which is easily solved when Q is p.d.
We will sketch the Active set method for convex QP

Primal Active set method for convex QP

- Primal active-set methods, at each iteration k, solves a quadratic subproblem in which the inequality constraints in the working set W_k are imposed as equalities
- the gradients a_i of the constraints in W_k are linearly independent
- first check whether $\boldsymbol{x}_k$ minimizes the quadratic $q(\boldsymbol{x})$ in W_k
- If not, compute a step p by solving a suitable equality-constrained QP subproblem on W_k
    - $p = \boldsymbol{x} - \boldsymbol{x}_k$, $g_k = Q\boldsymbol{x}_k + \boldsymbol{b}$
    - $p_k = \arg\min q(\boldsymbol{x}) = \arg\min q(\boldsymbol{x}_k + p) = \arg\min \frac{1}{2} p^T Qp + g_k^T p + \rho_k$ s.t. $a_i^T p = 0, i \in W_q$, where $\rho_k = \frac{1}{2}\boldsymbol{x}_k^T Q\boldsymbol{x}_k + \boldsymbol{b}^T \boldsymbol{x}_k$ is independent from p
    - for each $i \in W_k$, we have $a_i^T(\boldsymbol{x}_k + \alpha p_k) = a_i^T \boldsymbol{x}_k = d_i$ for all α. The constraints in W_k are also satisfied at $\boldsymbol{x}_k + \alpha p_k$.
    - if $p_k \neq \boldsymbol{0}$ then
        - $\alpha_k = \min\left(1, \min\{\frac{b_i - a_i^T x_k}{a_i^T p_k} : i \notin W_k, a_i^T p_k < 0\}\right)$;
        - $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k p_k$
    - else test KKT or update W_k

Active set method for convex QP

Active Set Method;

Choose a feasible $\mathbf{x}_0$;

Set W_0 to be a subset of the active constraints at $\mathbf{x}_0$;

for $k = 0, 1, 2, \ldots$;

$p_k = \arg\min\{\frac{1}{2}p^T G p + g_k^T p \quad s.t. \quad a_i^T p = 0, i \in W_k\}$;

if $p_k = 0$;

Compute λ_i s.t. $\sum_{i \in W_k} a_i \lambda_i = G\mathbf{x}_k + c$;

if $\lambda_i \geq 0$ for all $i \in W_k \cap I$

Stop $\mathbf{x}^* = \mathbf{x}_k$;

else

$j = \arg\min_{j \in W_k \cap I} \lambda_j$;

$\mathbf{x}_{k+1} = \mathbf{x}_k$; $W_{k+1} = W_k \backslash \{j\}$;

else * $p_k \neq 0$ *\

$\alpha_k = \min\left(1, \min\{\frac{b_i - a_i^T x_k}{a_i^T p_k} : i \notin W_k, a_i^T p_k < 0\}\right)$;

$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k p_k$;

if $\alpha_k < 1$;

$W_{k+1} = W_k \cup \{$ a blocking constraint $\}$;

else

$W_{k+1} = W_k$;

endfor

Quadratic penalty method

Transform a constrained problem into an unconstrained one.

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \dots, h; \end{array}$$

Penalty function

$$p(\boldsymbol{x}) = \sum_{j=1}^{h} h_j^2(\boldsymbol{x})$$

The quadratic penalty model is

$$\min q(\boldsymbol{x}) = f(\boldsymbol{x}) + \alpha \sum_{j=1}^{h} h_j^2(\boldsymbol{x}).$$

By driving α to ∞, we penalize the constraint violations with increasing severity. It makes good intuitive sense to consider a sequence of values $\{\alpha_k\}$ with $\alpha \to \infty$ as $k \to \infty$, and to seek the approximate minimizer $\boldsymbol{x}_k$ of $q(\boldsymbol{x}; \alpha_k)$ for each k. Because the penalty terms are smooth, we can use techniques from uncostrained optimization

Quadratic penalty method

$$\min q(\boldsymbol{x}) = f(\boldsymbol{x}) + \alpha \sum_{j=1}^{h} h_j^2(\boldsymbol{x}).$$

First and second order optimality conditions are that

$$\nabla q(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*) + 2\alpha \sum_{j=1}^{h} h_j(\boldsymbol{x}^*)\nabla h_j(\boldsymbol{x}^*) = \mathbf{0},$$

and that the hessian matrix of $q(\boldsymbol{x})$ in $\boldsymbol{x}^*$

$$\nabla^2 q(\boldsymbol{x}^*) = \nabla^2 f(\boldsymbol{x}^*) + 2\alpha \sum_{j=1}^{h} \left(h_j(\boldsymbol{x}^*)\nabla^2 h_j(\boldsymbol{x}^*) + \nabla h_j(\boldsymbol{x}^*)\nabla h_j(\boldsymbol{x}^*)^T\right)$$

is positive semidefinite

Quadratic penalty method

We can prove that when $\alpha \to \infty$ then $\boldsymbol{x}^*(\alpha) \to$ to a local minimum of the constrained problem, moreover

$$\lim_{\alpha \to \infty} 2\alpha h_j(\boldsymbol{x}^*(\alpha)) = \lambda_j^*$$

where λ_j^* is the optimal value of the lagrangean multiplier of the j-th constraint.

Quadratic penalty method

The hessian matrix of q is composed by two terms.

$$\nabla^2 q(\boldsymbol{x}^*) = \nabla^2 f(\boldsymbol{x}^*) + 2\alpha \sum_{j=1}^{h} \left(h_j(\boldsymbol{x}^*)\nabla^2 h_j(\boldsymbol{x}^*) + \nabla h_j(\boldsymbol{x}^*)\nabla h_j(\boldsymbol{x}^*)^T\right)$$

The first term is

$$\nabla^2 f(\boldsymbol{x}^*) + 2\alpha \sum_{j=1}^{h} h_j(\boldsymbol{x}^*)\nabla^2 h_j(\boldsymbol{x}^*)$$

which for $\alpha \to \infty$ becomes

$$\nabla^2 f(\boldsymbol{x}^*) + \sum_{j=1}^{h} \lambda_j^* \nabla^2 h_j(\boldsymbol{x}^*)$$

i.e. the hessian matrix of the Lagrangean function in $\boldsymbol{x}^*$

Quadratic penalty method

The hessian matrix of q is composed by two terms.

$$\nabla^2 q(\boldsymbol{x}^*) = \nabla^2 f(\boldsymbol{x}^*) + 2\alpha \sum_{j=1}^{h} \left(h_j(\boldsymbol{x}^*) \nabla^2 h_j(\boldsymbol{x}^*) + \nabla h_j(\boldsymbol{x}^*) \nabla h_j(\boldsymbol{x}^*)^T \right)$$

The second term is

$$\sum_{j=1}^{h} 2\alpha \nabla h_j(\boldsymbol{x}^*) \nabla h_j(\boldsymbol{x}^*)^T$$

whose norm diverge for $\alpha \to \infty$

From a practical viewpoint the matrix hessian becomes increasingly illconditioned as far as we converge to $\boldsymbol{x}^*$.

Barrier methods

Let consider an inequality constrained problem

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) & \leq 0 & j = 1, \ldots, k; \end{array}$$

We devide the feasible region into

- a frontier set $S_f := \{\boldsymbol{x} \in \mathbb{R}^n \,|\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}\}$ and
- a inner set $S_{int} := \{\boldsymbol{x} \in \mathbb{R}^n \,|\boldsymbol{g}(\boldsymbol{x}) < \boldsymbol{0}\}$

Barrier methods apply when $S_{int} \neq \emptyset$. They use a *barrier function* $v(\boldsymbol{x})$ which is continous in S_{int}, and s.t. $v(\boldsymbol{x}) \to \infty$ when $\boldsymbol{x} \to S_f$.

The model is

$$\min \; b(\boldsymbol{x}) = f(\boldsymbol{x}) + \alpha v(\boldsymbol{x}).$$

The logaritmic barrier model is

$$v(\boldsymbol{x}) = -\sum_{i=1}^{k} \log(-g_j(\boldsymbol{x}))$$

From a practical viewpoint the matrix hessian becomes increasingly illconditioned for increasing value of α.

Projected gradient method

Due to Rosen (1960, 1961). Let us start with linear equality constraints

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ t.c. \quad & A\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Start with a feasible solution $\boldsymbol{x}'$, $A\boldsymbol{x}' = \boldsymbol{b}$, and look for an improved solution $\boldsymbol{x} = \boldsymbol{x}' + \alpha \boldsymbol{d}$. Direction $\boldsymbol{d}$ must

- be normalized, i.e. $||\boldsymbol{d}|| = 1$
- satisfy $A(\boldsymbol{x}' + \alpha \boldsymbol{d}) - \boldsymbol{b} = \boldsymbol{0}$, which is $A\boldsymbol{d} = \boldsymbol{0}$
- minimize the directional derivative $\nabla f(\boldsymbol{x}')^T \boldsymbol{d}$ in $\boldsymbol{x}'$

Projected gradient method

This leads to

$$\begin{array}{ll} \min & \nabla f(\boldsymbol{x}')^T \boldsymbol{d} \\ t.c. & 1 - \boldsymbol{d}^T \boldsymbol{d} = 0 \\ & A\boldsymbol{d} = \boldsymbol{0} \end{array}$$

The lagrangean function is

$$L(\boldsymbol{d}, \boldsymbol{\lambda}, \lambda_0) = \nabla f(\boldsymbol{x}')^T \boldsymbol{d} + \boldsymbol{\lambda}^T A\boldsymbol{d} + \lambda_0 (1 - \boldsymbol{d}^T \boldsymbol{d})$$

and by imposing the necessary optimality conditions

$$\begin{array}{rcl} \nabla_{\boldsymbol{d}} L & = & \nabla f(\boldsymbol{x}') + \boldsymbol{\lambda}^T A - 2\lambda_0 \boldsymbol{d} = \boldsymbol{0} \\ \nabla_{\boldsymbol{\lambda}} L & = & A\boldsymbol{d} = \boldsymbol{0} \\ \nabla_{\lambda_0} L & = & (1 - \boldsymbol{d}^T \boldsymbol{d}) = 0 \end{array}$$

you (try as an exercise) obtain

$$\boldsymbol{d} = -\frac{\left(I - A^T (AA^T)^{-1} A\right) \nabla f(\boldsymbol{x}')}{\left\| \left(I - A^T (AA^T)^{-1} A\right) \nabla f(\boldsymbol{x}') \right\|}$$

Projected gradient method

$$\boldsymbol{d} = -\frac{\left(I - A^T(AA^T)^{-1}A\right)\nabla f(\boldsymbol{x}')}{\|\left(I - A^T(AA^T)^{-1}A\right)\nabla f(\boldsymbol{x}')\|}$$

- $-\nabla f(\boldsymbol{x}')$ is the most improving direction of $f(\boldsymbol{x})$ in $\boldsymbol{x}'$
- $\boldsymbol{d}$ is the projection of $-\nabla f(\boldsymbol{x}')$ into the hyperplane $A\boldsymbol{x} = \boldsymbol{b}$.
- The matrix $P = \left(I - A^T(AA^T)^{-1}A\right)$ is called *projection matrix*
- In practice, $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha\boldsymbol{d}$, you use $\boldsymbol{d} = -P\nabla f(\boldsymbol{x}')$, and you determine α with, e.g. Armijo

Projected gradient method

In a problem with generic equality constraints

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \dots, h \end{array}$$

we use Taylor for obtain linear constraints nearby the current feasible solution $\mathbf{x}'$

$$h_j(\mathbf{x}) = h_j(\mathbf{x}') + \nabla h_j(x')^T(\mathbf{x} - \mathbf{x}'),$$

hence

$$\nabla h_j(\mathbf{x}')^T \mathbf{x} - \nabla h_j(\mathbf{x}')^T \mathbf{x}' = 0, \quad j = 1, \dots, h.$$

By setting $A = \left[\frac{\partial h(\mathbf{x}')}{\partial x}\right]^T$, and $\mathbf{b} = \left[\frac{\partial h(\mathbf{x}')}{\partial x}\right]^T \mathbf{x}'$, we obtain the following linear constrained model

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ t.c. & A\mathbf{x} = \mathbf{b} \end{array}$$

The projection matrix $P(\mathbf{x}') = (I - A^T(AA^T)^{-1}A)$, depends from $\mathbf{x}'$ through the matrix A, and we use $\mathbf{d} = -P(\mathbf{x}')\nabla f(\mathbf{x}')$.

Projected gradient method

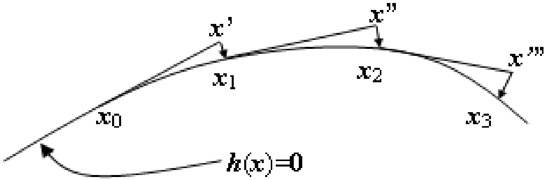
Since $\boldsymbol{x}_k = \boldsymbol{x}'$, for every value of $\alpha > 0$, the new point $\boldsymbol{x}'' = \boldsymbol{x}_k + \alpha \boldsymbol{d}$, likely does not satisfy the original nonlinear equality constraints, $\boldsymbol{h}(\boldsymbol{x}'') \neq \boldsymbol{0}$, we need to apply a corrective step $\boldsymbol{x}'' \to \boldsymbol{x}_{k+1}$.

By imposing

$$P(\boldsymbol{x}_k)(\boldsymbol{x}_{k+1} - \boldsymbol{x}'') = \boldsymbol{0},$$

and $\boldsymbol{h}(\boldsymbol{x}_{k+1}) = \boldsymbol{0}$ we obtain

$$\boldsymbol{x}_{k+1} \approx \boldsymbol{x}'' - A^T(AA^T)^{-1}\boldsymbol{h}(\boldsymbol{x}'').$$



The corrective step is applied till $\boldsymbol{h}(\boldsymbol{x}_{k+1})$ is small enough while the whole algorithm stops when $P(\boldsymbol{x}')\nabla f(\boldsymbol{x}') \approx \boldsymbol{0}$.

Augmented lagrangean method

This approach (Bertsekas 1976) combines the use of the langrangean function with the quadratic penalty functions. The idea is that of approximating the lagrangean multipliers.
In a generic problem with equality constraints

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \dots, h \end{aligned}$$

We introduce the **augmented langrangean** function:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \rho) = f(\boldsymbol{x}) + \sum_{j=1}^{h} \lambda_j h_j(\boldsymbol{x}) + \rho \sum_{j=1}^{h} h_j^2(\boldsymbol{x})$$

When $\lambda_j = 0$ we have the penalty function
Moreover if we know λ_j^* for each $\rho > 0$ by minimizing $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \rho)$ with respect to $\boldsymbol{x}$ we get $\boldsymbol{x}^*$ (Fletcher 1987)
If $\boldsymbol{\lambda}^k$ is a valid approximation of $\boldsymbol{\lambda}^*$, then we can approximate $\boldsymbol{x}^*$ by minimizing $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^k, \rho)$ even for small values of ρ

ρ must guarantee that $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^k, \rho)$ has a local minimum with respect to $\boldsymbol{x}$ and not just a stationary point

Augmented lagrangean method

To understand this technique it suffices to compare the stationary conditions of L and $\mathcal{L}$ in $\boldsymbol{x}^*$.
For $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{h} (\lambda_j^k + 2\rho h_j) \frac{\partial h_j}{\partial x_i} = 0, \quad i = 1, \ldots, n.$$

For L :

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{h} \lambda_j^k \frac{\partial h_j}{\partial x_i} = 0, \quad i = 1, \ldots, n.$$

We see that when the minimum of $\mathcal{L}$ approaches to $\boldsymbol{x}^*$, then:

$$\lambda_j^k + 2\rho h_j \to \lambda_j^*$$

Augmented lagrangean method

This lead to the following algorithm

- Set $k = 0$; initialize $\boldsymbol{\lambda}^k$ and ρ;
- While $||\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^k, \rho)|| > \varepsilon$ do
  - compute $\boldsymbol{x}_k^*$ by solving $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^k, \rho)$ with respect to $\boldsymbol{x}$, with your preferred approach for unconstrained optimization
  - update $\boldsymbol{\lambda}$ with
  $$\lambda_j^{k+1} := \lambda_j^k + 2\rho h_j(\boldsymbol{x}_k^*)$$
  - Eventualy update ρ.

SQP (Sequential Quadratic Programming)

The idea: apply Newton's method for finding $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ from the KKT conditions of constrained problem.
Each Newton step can be reduced to the solution of a QP.
Let us consider the general problem

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_i(\boldsymbol{x}) & \leq 0 & i = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array} \qquad (7)$$

and its lagrangean model

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j g_j(\boldsymbol{x}) + \sum_{j=1}^{h} \mu_j h_j(\boldsymbol{x})$$

We are given an approximation $(\boldsymbol{x}_k, \boldsymbol{\lambda}_k, \boldsymbol{\mu}_k)$, with $\boldsymbol{\lambda}_k \geq 0$, $k = 1, 2, \ldots,$ of the solution and of the lagrangean multipliers and we know the hessian matrix of L

$$\nabla^2 L(\boldsymbol{x}_k) = H(\boldsymbol{x}_k) + \sum_{j=1}^{k} \lambda_j^k \nabla^2 g_j(\boldsymbol{x}_k) + \sum_{j=1}^{h} \mu_j^k \nabla^2 h_j(\boldsymbol{x}_k).$$

SQP (Sequential Quadratic Programming)

We can prove that the Newton direction $\boldsymbol{d}$ for computing $\boldsymbol{x}_{k+1}$ from $\boldsymbol{x}_k$,

$$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \boldsymbol{d}_k$$

can be obtained by solving the following QP with equality and inequality constraints:

$$\begin{aligned}
\min\ \phi(\boldsymbol{d}) &= f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \nabla^2 L(\boldsymbol{x}_k)\boldsymbol{d} \\
\boldsymbol{g}(\boldsymbol{x}_k) &+ \left[\frac{\partial \boldsymbol{g}(\boldsymbol{x}_k)}{\partial \boldsymbol{x}}\right]^T \boldsymbol{d} \leqslant \boldsymbol{0}, \\
\boldsymbol{h}(\boldsymbol{x}_k) &+ \left[\frac{\partial \boldsymbol{h}(\boldsymbol{x}_k)}{\partial \boldsymbol{x}}\right]^T \boldsymbol{d} = \boldsymbol{0}
\end{aligned}$$

By solving the QP model we get, besides $\boldsymbol{d}$ and $\boldsymbol{x}_{k+1}$, also $\boldsymbol{\lambda}_{k+1}$ and $\boldsymbol{\mu}_{k+1}$. So we have all the data for the next iteration
The stopping criterion is on a threshold on the norm of $\boldsymbol{d}$

SQP (Sequential Quadratic Programming)

The SQP method returns a point which satisfies KKT conditions. Hence all not regular points (those which do not satisfy constraints qualification) are missed by the algorithm.