Nonlinear Optimization

Constrained problems

Niclas Börlin

Department of Computing Science

Umeå University
niclas.borlin@cs.umu.se

December 13, 2007

Optimality conditions for constrained problems

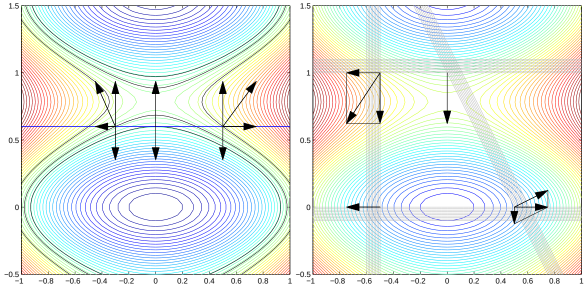
- A minimizer x^* to a minimization problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \mathcal{E} \\ & c_i(x) \geq 0, \quad i \in \mathcal{I} \end{aligned}$$

must satisfy

$$p^T \nabla f(x^*) \geq 0$$

for all feasible directions p.



- Consider a problem with linear equality constraints, i.e.

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

where A is assumed to have full rank.
- The constrained problem may be rewritten to the unconstrained problem

$$\min_{v \in \Re^r} \phi(v) = f(\bar{x} + Zv),$$

where $\bar{x}$ is a feasible point and $Z \in \Re^{n \times r}$ is a basis for $\mathcal{N}(A)$.
- The function $\phi(v)$ is called the *reduced function*.

Necessary conditions for a minimizer

- The necessary conditions for the reduced problem is

$$\begin{aligned} \nabla \phi(v) &= Z^T \nabla f(\bar{x} + Zv) = Z^T \nabla f(x) = 0 \text{ and} \\ \nabla^2 \phi(v) &= Z^T \nabla^2 f(\bar{x} + Zv) Z = Z^T \nabla^2 f(x) Z \text{ positive semidefinite,} \end{aligned}$$

where $x = \bar{x} + Zv$.
- The expression $Z^T \nabla f(x)$ is called the *reduced gradient* and $Z^T \nabla^2 f(x) Z$ the *reduced Hessian*.
- If the null space matrix Z is an orthogonal projection matrix, they are called *projected* gradient and Hessian, respectively.

- The second order condition corresponds to

$$v^T Z^T \nabla^2 f(x^*) Z v \geq 0 \quad \forall v,$$

  which may be rewritten as

$$p^T \nabla^2 f(x^*) p \geq 0 \quad \forall p \in \mathcal{N}(A),$$

  where $p = Zv$.
- Thus, the Hessian in x^* has to be positive semidefinite on the null space of A.
- **This does not mean that the Hessian in x^* has to be positive semidefinite on the whole $\Re^n$.**

Sufficient conditions for linear equality constraints

- If x^* satisfies
  - $Ax^* = b$,
  - $Z^T \nabla f(x^*) = 0$, and
  - $Z^T \nabla^2 f(x^*) Z$ is positive definite,

  where Z is a basis matrix for $\mathcal{N}(A)$, then x^* is a strict local minimizer of f over $\{x : Ax = b\}$.

Example

- Consider the problem

$$\min_x f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3,$$
$$\text{s.t. } x_1 - x_2 + 2x_3 = 2,$$

  with gradient and Hessian functions

$$\nabla f(x) = \begin{bmatrix} 2x_1 - 2 \\ 2x_2 \\ -2x_3 + 4 \end{bmatrix} \text{ and } \nabla^2 f(x) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

- As null space matrix of the constraint matrix

$$A = [1, -1, 2]$$

  we may choose

$$Z = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

- In the feasible point $x = [2, 0, 0]^T$, the reduced gradient is

$$Z^T \nabla f(x) = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}^T \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

  and x is not a local minimizer.
- In the feasible point $x^* = [2.5, -1.5, -1]^T$, the gradient of f is $[3, -3, 6]^T$ and $Z^T \nabla f(x) = 0$.
- The reduced Hessian is

$$Z^T \nabla^2 f(x^*) Z = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}^T \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ -4 & 6 \end{bmatrix}$$

  and is positive definite.
- Thus, the second order sufficient conditions are satisfied and x^* is a strict local minimizer of f.
- Notice that $\nabla^2 f(x)$ itself is not positive definite.

The Lagrange multipliers

- Let x^* be a minimizers and $Z \in \Re^{n\times r}$ a null space matrix for A.
- The gradient $\nabla f(x^*)$ may be expressed as the sum of its null space and range space components, i.e.

$$\nabla f(x^*) \quad = \quad Zv^* + A^T\lambda^*,$$

  where $v^* \in \Re^r$ and $\lambda^* \in \Re^m$.
- Together with the first order conditions we get

$$Z^T\nabla f(x^*) = Z^TZv^* + Z^TA^T\lambda^* = Z^TZv^* + (\underbrace{AZ}_{=0})^T\lambda^* = Z^TZv^* = 0$$

$$\Downarrow$$

$$Zv^* = 0.$$

- Thus, a minimizer satisfies

$$\nabla f(x^*) = A^T\lambda^*,$$

  i.e. in a local minimum, the gradient of the objective function is a linear combination of the gradients of the constraints.

- The coefficients in the vector λ^* are called *Lagrange multipliers*.
- The constraint and first order condition may be formulated in one system equation of $n+m$ equations and $n+m$ unknowns in x and λ:

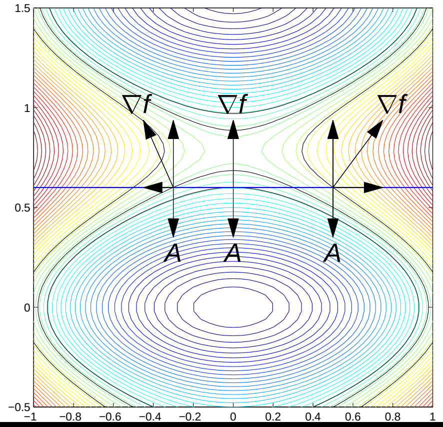
$$\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} \nabla f(x) \\ b \end{bmatrix}$$

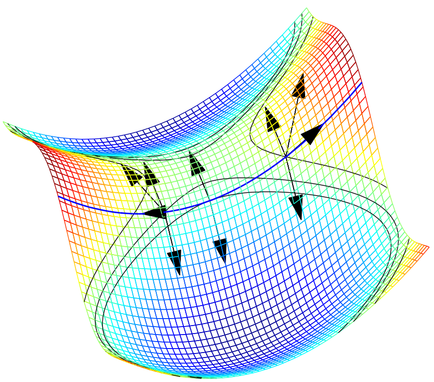
Example

$$\begin{array}{ll} \min_x & f(x) = x^2 + \sin^2 2y \\ \text{s.t.} & -y = -0.6 \end{array}$$

$$A = \begin{bmatrix} 0 & -1 \end{bmatrix}, Z = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, x^* = \begin{bmatrix} 0 \\ 0.6 \end{bmatrix},$$

$$\nabla^2 f(x^*) = \begin{bmatrix} 2 & 0 \\ 0 & -5.9 \end{bmatrix}, Z^T\nabla^2 f(x^*)Z = \begin{bmatrix} 2 \end{bmatrix}.$$





Lagrange multipliers and sensitivity

- The Lagrange multipliers may be used to estimate the sensitivity of the min value $f(x^*)$ with respect to the constraints.
- Assume we have found a solution x^* to

$$\min f(x) \text{ s.t. } Ax = b.$$

- Consider the perturbed constraints $Ax = b + \delta$.
- If the perturbation δ is small enough, the solution $\bar{x}$ to the perturbed problem will be close to x^* and

$$\begin{aligned} f(\bar{x}) &\approx f(x^*) + (\bar{x} - x^*)^T\nabla f(x^*) = f(x^*) + (\bar{x} - x^*)^TA^T\lambda^* \\ &= f(x^*) + (A\bar{x} - Ax^*)^T\lambda^* = f(x^*) + (b + \delta - b)^T\lambda^* \\ &= f(x^*) + \delta^T\lambda^* = f(x^*) + \sum_{i=1}^m \delta_i\lambda_{*i}. \end{aligned}$$

- Thus, if element i of the right hand side of the constraint is modified by δ_i, the optimal objective value will change with about $\delta_i\lambda_{*i}$.
- For this reason, the Lagrange multipliers are sometimes called *shadow prices* or *dual variables*.

The Lagrangian function

- Define the *Lagrangian function* of x and λ as
$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i=1}^{m} \lambda_i (a_i^T x - b_i) = f(x) - \lambda^T (Ax - b).$$
- The gradients of the Lagrangian are
$$\nabla_x \mathcal{L}(x,\lambda) = \nabla f(x) - A^T \lambda$$
and
$$\nabla_\lambda \mathcal{L}(x,\lambda) = b - Ax.$$
- The first order condition on the Lagrangian
$$\nabla \mathcal{L}(x^*,\lambda^*) = 0$$
correspond to the first order condition on the constrained problem.
- A local minimizer to the constrained problem is a stationary point to the Lagrangian.

Optimality conditions for linear inequality constrained problems

- Consider a problem with linear inequality constraints, i.e.
$$\min_x f(x) \text{ s.t. } Ax \geq b,$$
where A is assumed to have full rank.
- The active constraint in a point x^* will determine if x^* is a minimizer.
- Our problem may thus be rewritten as
$$\min_x f(x) \text{ s.t. } \hat{A}x = \hat{b},$$
where $\hat{A}$ and $\hat{b}$ contains the active constraints.
- If Z is a null space matrix $\hat{A}$, the first order condition becomes
$$Z^T \nabla f(x^*) = 0 \text{ or } \nabla f(x^*) = \hat{A}^T \hat{\lambda}_*,$$
where $\hat{\lambda}_* \geq 0$ has the Lagrange multipliers for the active constraints.
- The second order necessary condition is that $Z^T \nabla^2 f(x^*) Z$ must be positive semidefinite.

Complementary slackness

- If we define the Lagrange multiplier of an inactive constraint to be zero, we may describe the inequality conditions as
$$\lambda_{*i}(a_i^T x^* - b_i) = 0, \quad i = 1, \ldots, m.$$
- This condition is called *complementary slackness* and means that either the constraint is active ($a_i^T x^* - b_i = 0$) or the Lagrange multiplier is zero ($\lambda_{*i} = 0$).
- At least one of the two must be true.
- The case when both cannot be true at the same time is called *strict complementarity*.
- Without strict complementarity, a Lagrange multiplier may be zero even if the constraint is active.
- In such a case, the constraint is called *degenerate*.

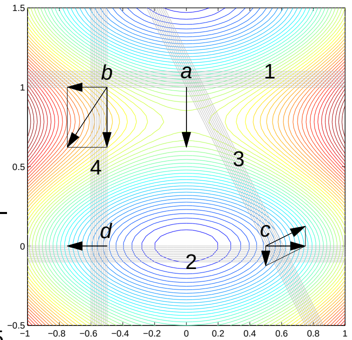
Complementary slackness

For the problem

$$\begin{aligned} \min_x \quad & f(x) = x^2 + \sin^2 2y \\ \text{s.t.} \quad & \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ -2 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq \begin{bmatrix} -1 \\ 0 \\ -1 \\ -1 \end{bmatrix} \end{aligned}$$

there are four corners, two of which are degenerate.

Point	Active constraints	(x,y)	∇f	λ
a	1,3	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -1.5 \end{bmatrix}$	$1.5, 0$
b	1,4	$\begin{bmatrix} -0.5 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -1 \\ -1.5 \end{bmatrix}$	$1.5, -0.5$
c	2,3	$\begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$-0.5, -0.5$
d	2,4	$\begin{bmatrix} -0.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 0 \end{bmatrix}$	$0, -0.5$



Necessary condition, linear inequalities

- In summary, the following has to be satisfied for a point x^* to be a minimizer of a function f on the set $\{x : Ax \geq b\}$:
  - $Ax^* \geq b$,
  - $\nabla f(x^*) = A^T\lambda^* \Leftrightarrow Z^T\nabla f(x^*) = 0$,
  - $\lambda^* \geq 0$,
  - ${\lambda^*}^T(Ax^* - b) = 0$ and
  - $Z^T\nabla^2 f(x^*)Z$ positive semidefinite,

  for some vector λ^* of Lagrange multipliers and where Z is a null space matrix for the matrix $\hat{A}$ of the active constraints in x^*.

Sufficient conditions, linear inequalities I

- If we have strict complementarity we may extend to sufficient conditions in a straightforward manner:
- If
  - $Ax^* \geq b$,
  - $\nabla f(x^*) = A^T\lambda^*$,
  - $\lambda^* \geq 0$,
  - we have strict complementarity, and
  - $Z^T\nabla^2 f(x^*)Z$ positive definite,

  then x^* is a strict local minimizer of f on the set $\{x : Ax \geq b\}$.

Why strict complementarity is needed

- The point x^* is also a strict local minimizer on the set $\{x : \hat{A}x = \hat{b}\}$, i.e. f increases in all directions p such that $\hat{A}p = 0$:
- Study a direction p such that $\hat{A}p \geq 0$, where some component is strictly positive, i.e. p points into the feasible set.
- Since $\nabla f(x^*) = A^T\lambda^* = \hat{A}^T\hat{\lambda}_*$, then

$$p^T\nabla f(x^*) = p^T\hat{A}^T\hat{\lambda}_* = (\hat{A}p)^T\hat{\lambda}_*.$$

- With strict complementarity, $(\hat{A}p)^T\hat{\lambda}_* > 0$, i.e. p is an ascent direction and x^* is a strict minimizer.
- Without strict complementarity, $(\hat{A}p)^T\hat{\lambda}_* = 0$ may be true for some p, which means we cannot tell anything about x^* with only first order information.
- However, if we drop the degenerate constraints, we may formulate sufficient conditions on the remaining constraints.

Sufficient conditions, linear inequalities II

- Let $\hat{A}_+$ contain the rows of $\hat{A}$ corresponding to the non-degenerate constraints in x^*.
- Let Z_+ be a null space matrix to $\hat{A}_+$.
- If x^* satisfies
  - $Ax^* \geq b$,
  - $\nabla f(x^*) = A^T\lambda^*$,
  - $\lambda^* \geq 0$,
  - ${\lambda^*}^T(Ax^* - b) = 0$, and
  - $Z_+^T\nabla^2 f(x^*)Z_+$ positive definite,

  then x^* is a strict local minimizer to the inequality constrained problem.

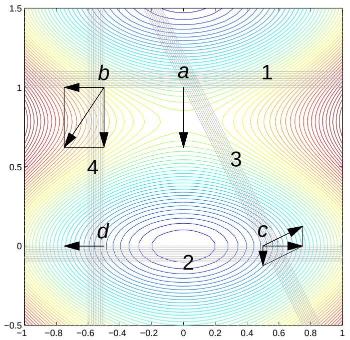
Example

For the problem

$$\min_x \quad f(x) = x^2 + \sin^2 2y$$

$$\text{s.t.} \quad \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ -2 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq \begin{bmatrix} -1 \\ 0 \\ -1 \\ -1 \end{bmatrix}.$$

	λ	$\nabla^2 f$	Z_+	$Z_+^T \nabla^2 f Z_+$
a	$1.5, 0$	$\begin{bmatrix} 2 & 0 \\ 0 & -5.2 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$[2]$
b	$1.5, -0.5$	$\begin{bmatrix} 2 & 0 \\ 0 & -5.2 \end{bmatrix}$	$[]$	$[]$
c	$-0.5, -0.5$	$\begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$	$[]$	$[]$
d	$0, -0.5$	$\begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$[8]$



Optimality conditions for non-linear constraints

- Non-linear optimization problems with non-linear are formulated as
$$\min_x f(x) \text{ s.t. } c_i(x) = 0, i = 1, \ldots, m$$
for equality constraints, and
$$\min_x f(x) \text{ s.t. } c_i(x) \geq 0, i = 1, \ldots, m$$
for inequality constraints.
- We will assume that the solution point x^* is *regular*, i.e. that the gradients of the active constraints in x^* $\{\nabla c_i(x^*) : c_i(x^*) = 0\}$ are linearly independent.
- The optimality conditions are expressed in terms of the Lagrangian function
$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x) = f(x) - \lambda^T c(x),$$
where λ is a vector of Lagrange multipliers and c is a vector of constraint functions $\{c_i\}$.

Necessary conditions for equality constraints

- Let x^* be a local minimizer for f under the constraints $c(x) = 0$ and $Z(x^*)$ be a null space matrix for the Jacobian $\nabla c(x^*)^T$ of the constraints.
- If x^* is a regular point, then there exists a Lagrangian vector λ^* such that
  - $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \Leftrightarrow Z(x^*)^T \nabla f(x^*) = 0$ and
  - $Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*)$ positive semi-definite.

Sufficient conditions for equality constraints

- Let x^* be a point such that $c(x^*) = 0$ and $Z(x^*)$ is a basis for the null space of $\nabla c(x^*)^T$.
- If there exists a vector λ^* such that
  - $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$ and
  - $Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*)$ is positive definite,

  then x^* is a strict local minimizer to f on the constraint set $\{x : c(x) = 0\}$.

Linear constraints revisited

- For linear constraints $c(x) = Ax - b$, the Jacobian is $\nabla c(x)^T = A$ and the first order conditions

$$Z(x^*)^T \nabla f(x^*) = 0 \quad \Leftrightarrow \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \nabla c(x^*)\lambda = 0$$

becomes

$$Z^T \nabla f(x^*) = 0 \quad \Leftrightarrow \quad \nabla f(x^*) = A^T \lambda^*.$$

- The second order necessary (sufficient) conditions becomes that

$$\nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) = \nabla^2 f(x^*)$$

should be positive semi-definite (positive definite).

Necessary conditions for inequality constraints

- Let x^* be a local minimizer for f under the constraints $c(x) \geq 0$ and $Z(x^*)$ be a null space matrix for the Jacobian of the active constraints in x^*.
- If x^* is a regular point, then there exists a Lagrangian vector λ^* such that
  - $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \Leftrightarrow Z(x^*)^T \nabla f(x^*) = 0$,
  - $\lambda^* \geq 0$,
  - $\lambda^{*T} c(x^*) = 0$, and
  - $Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*)$ positive semi-definite.
- The condition $\lambda^{*T} c(x^*) = 0$ is the non-linear version of the complementary slackness condition.

Sufficient conditions for inequality constraints

- Let x^* be a points such that $c(x^*) \geq 0$. If there exists a vector λ^* such that
  - $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$,
  - $\lambda^* \geq 0$,
  - $\lambda^{*T} c(x^*) = 0$, and
  - $Z_+(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z_+(x^*)$ is positive definite,

  where $Z_+(x^*)$ is a basis for the null space of the Jacobian of the non-degenerate constraints in x^*, then x^* is a strict local minimizer to f on the constraint set $\{x : c(x) \geq 0\}$.
- **The necessary and sufficient conditions for the non-linear inequality constraints are often called the KKT conditions (Karush-Kuhn-Tucker conditions).**

Duality

The concept of duality is that for each minimization problem, there is a corresponding maximization problem that under some circumstances both problems have the same optimum.
Define

$$\mathcal{F}^*(x) = \max_{y \in Y} \mathcal{F}(x, y)$$

and

$$\mathcal{F}_*(y) = \min_{x \in X} \mathcal{F}(x, y).$$

The problem

$$\min_{x \in X} \mathcal{F}^*(x) = \min_{x \in X} \max_{y \in Y} \mathcal{F}(x, y)$$

is called a min-max problem and the problem

$$\max_{y \in Y} \mathcal{F}_*(y) = \max_{y \in Y} \min_{x \in X} \mathcal{F}(x, y)$$

is called a max-min problem. These problems are each others *duals*. The min-max problem is called the *primal* problem and $\mathcal{F}^*(x)$ is called the *primal function*. The max-min problem is called the *dual* problem and $\mathcal{F}_*(y)$ is called the *dual function*.

Weak and strong duality

Each $x \in X$ and $y \in Y$ satisfies

$$\mathcal{F}_*(y) = \min_{x \in X} \mathcal{F}(x,y) \le \mathcal{F}(x,y) \le \max_{y \in Y} \mathcal{F}(x,y) = \mathcal{F}^*(x)$$

or

$$\mathcal{F}_*(y) \le \mathcal{F}^*(x).$$

This is called *weak duality*. A consequence of weak duality is that the primal problem is bounded from below by $\mathcal{F}_*(y)$.
A point (x^*, y_*) satisfies the *saddle-point condition* for $\mathcal{F}$ if

$$\mathcal{F}(x^*, y) \le \mathcal{F}(x^*, y_*) \le \mathcal{F}(x, y_*)$$

for all $x \in X$ and $y \in Y$.
If there exists a point (x^*, y_*) that satisfies the saddle-point condition, then the solution value of the primal and the dual problem is the same, i.e.

$$\min_{x \in X} \max_{y \in Y} \mathcal{F}(x,y) = \max_{y \in Y} \min_{x \in X} \mathcal{F}(x,y).$$

This is called *strong duality*.

Duality and the Lagrange multipliers

Study the non-linear problem

$$\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & c_i(x) \ge 0, \quad i = 1, \dots, m \end{array}$$

and its corresponding Lagrangian function

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x),$$

where $x \in \Re^n$, $\lambda \in \Re^m$, $\lambda \ge 0$. Define the primal function

$$\mathcal{L}^*(x) = \max_{\lambda \ge 0} \mathcal{L}(x, \lambda)$$

and study $\mathcal{L}^*(x)$ for a fixed x:

$$\mathcal{L}^*(x) = \max_{\lambda \ge 0} \left(f(x) - \lambda^T c(x) \right).$$

For a feasible point, $c(x) \ge 0$ and $\mathcal{L}^*(x) = f(x)$. For an infeasible point, some constraint $c_i(x)$ will be negative and $\mathcal{L}^*(x)$ will be without bound.
If we formulate the primal problem as

$$\min_x \mathcal{L}^*(x),$$

then it will be the same as our original constrained problem.

Duality and the Lagrange multipliers

We may use min-max-duality to formulate the dual problem. For each $\lambda \ge 0$, define the dual function

$$\mathcal{L}_*(\lambda) = \min_x \mathcal{L}(x, \lambda)$$

and the dual max-min-problem

$$\max_{\lambda \ge 0} \mathcal{L}_*(\lambda).$$

Some methods try to solve the dual problem instead of the primal.