Introduction to constrained optimization

Gilles Gasso

INSA Rouen - ASI Departement
Laboratory LITIS

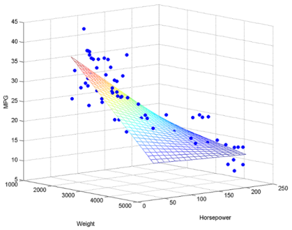
September 19, 2019

Plan

Constrained optimizing problems

- Where do these problems come from?
- What is the connection with Machine Learning?
- How to formulate them ?

Example 1: sparse Regression



- Output to be predicted: $y \in \mathbb{R}$
- Input variables: $\mathbf{x} \in \mathbb{R}^d$
- Linear model: $f(x) = \mathbf{x}^\top \boldsymbol{\theta}$
- $\boldsymbol{\theta} \in \mathbb{R}^d$: parameters of the model

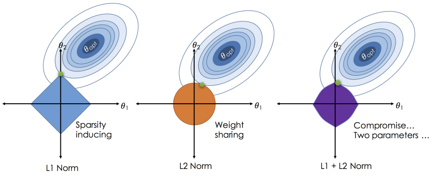
Determination of a sparse $\boldsymbol{\theta}$

- Minimization of square error
- Only a few paramters are non-zero

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^d} \frac{1}{2} \sum_{i=1}^N (y_i - \mathbf{x}_i^\top \boldsymbol{\theta})^2$$

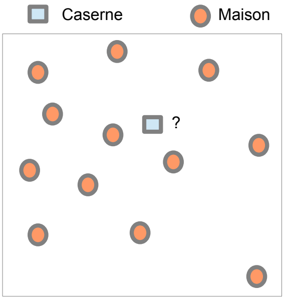
$$\text{s.c. } \|\boldsymbol{\theta}\|_p \leq k$$

with $\|\boldsymbol{\theta}\|_p^p = \sum_{j=1}^d |\theta_j|^p$



http://www.ds100.org/sp17/assets/notebooks/linear_regression/Regularization.html

Example 2: where to settle the firehouse?



- House M_i: defined by its coordinates $\mathbf{z}_i = [x_i \;\; y_i]^\top$
- Let $\boldsymbol{\theta}$ be the coordinates of the firehouse
- Minimize the distance from the firehouse to the farthest house

Problem formulation

$$\min_{\boldsymbol{\theta}} \max_{i=1,\cdots,N} \|\boldsymbol{\theta} - \mathbf{z}_i\|^2$$

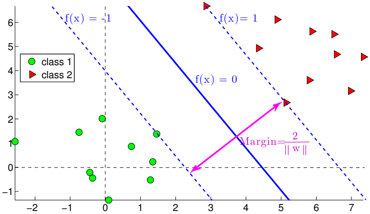
Equivalent problem

$$\min_{t \in \mathbb{R}, \boldsymbol{\theta} \in \mathbb{R}^2} t$$
$$\text{s.c. } \|\boldsymbol{\theta} - \mathbf{z}_i\|^2 \leq t \quad \forall\, i = 1, \cdots, N$$

Third example: Support vector machine (SVM)

- $\mathcal{D} = \{(x_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}\}_{i=1}^N$ linearly separable points
- Goal: find a function $f(\mathbf{x}) = \boldsymbol{\theta}^\top \mathbf{x} + b$ that correctly predicts the class of each $\mathbf{x}_i$ while maximizing the margin between positives and negatives

$$\begin{array}{lll} \min_{\boldsymbol{\theta} \in \mathbb{R}^d, b \in \mathbb{R}} & \frac{1}{2}\|\boldsymbol{\theta}\|^2 & \text{maximisation of the margin} \\ \text{s.t.} & y_i(\boldsymbol{\theta}^\top \mathbf{x}_i + b) \geq 1 \quad \forall i = 1, \cdots, N & \text{correct classification} \end{array}$$



SVM example: where are the constraints?

$$\begin{array}{lll} \min_{\boldsymbol{\theta},b} & \frac{1}{2}\|\boldsymbol{\theta}\|^2 & \text{function } J(\boldsymbol{\theta}) \text{ to be minimized} \\ \text{s.t.} & y_i(\boldsymbol{\theta}^\top \mathbf{x}_i + b) \geq 1 \quad \forall i = 1, \cdots, N & \text{Contraints to be satisfied} \end{array}$$

Contraints

- How many constraints ?
  - N constraints $y_i(\boldsymbol{\theta}^\top \mathbf{x}_i + b) \geq 1$ with $i = 1, \cdots, N$
- Type of constraints ?
  - Inequality constraints

Goal

Find the minimum $\boldsymbol{\theta}^*$ of $J(\boldsymbol{\theta})$ such as every constraint being satisfied

Constrained optimization

Elements of the problem

- $\boldsymbol{\theta} \in \mathbb{R}^d$: vector of unknown real parameters
- $J : \mathbb{R}^d \rightarrow \mathbb{R}$: the function to be minimized on its domain dom$J$
- f_i and g_j are differentiable functions of $\mathbb{R}^d$ on $\mathbb{R}$

Formulation of the primal problem $\mathcal{P}$

$$\begin{array}{llll} \min_{\boldsymbol{\theta} \in \mathbb{R}^d} & J(\boldsymbol{\theta}) & & \text{function to be minimized} \\ \text{s.c.} & f_i(\boldsymbol{\theta}) = 0 & \forall i = 1, \cdots, n & n \;\; \text{equality Constraints} \\ & g_j(\boldsymbol{\theta}) \leq 0 & \forall j = 1, \cdots, m & m \;\; \text{inequality Constraints} \end{array}$$

Feasibility

Let $p^* = \min_\theta \{J(\theta) \text{ such that } f_i(\theta) = 0 \; \forall i \text{ and } g_j(\theta) \leq 0 \, \forall j\}$

- If $p^* = \infty$ then the problem does not admit a feasible solution

Characterization of the solutions

Feasibility domain

The feasible domain is defined by the set of constraints

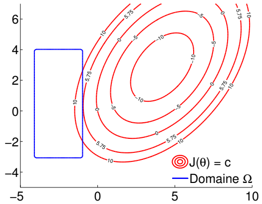
$$\Omega(\boldsymbol{\theta}) = \left\{\boldsymbol{\theta} \in \mathbb{R}^d;\ \ f_i(\boldsymbol{\theta}) = 0\ \forall i \text{ and } g_j(\boldsymbol{\theta}) \leq 0\,\forall j\right\}$$

Feasible points

- *$\boldsymbol{\theta}_0$ is feasible if $\boldsymbol{\theta}_0 \in domJ$ and $\boldsymbol{\theta}_0 \in \Omega(\boldsymbol{\theta})$ ie θ_0 fulfills all the constraints and $J(\theta_0)$ has a finite value*
- *$\boldsymbol{\theta}^*$ is a global solution of the problem if $\boldsymbol{\theta}^*$ is a feasible solution such that $J(\theta^*) \leq J(\boldsymbol{\theta})$ for every $\boldsymbol{\theta}$*
- *$\hat{\boldsymbol{\theta}}$ is a local optimal solution if $\hat{\boldsymbol{\theta}}$ is feasible and $J(\hat{\boldsymbol{\theta}}) \leq J(\boldsymbol{\theta})$ for every $\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\| \leq \epsilon$*

Example 1

$$\min_{\theta} \quad 0.9\theta_1^2 - 0.74\theta_1\theta_2$$
$$+0.75\theta_1^2 - 5.4\theta_1 - 1.2\theta_2$$
$$\text{s.c.} \quad -4 \leq \theta_1 \leq -1$$
$$-3 \leq \theta_2 \leq 4$$



- Parameters: $\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$
- Objective function:

$$J(\boldsymbol{\theta}) = 0.9\theta_1^2 - 0.74\theta_1\theta_2 + 0.75\theta_1^2 - 5.4\theta_1 - 1.2\theta_2$$

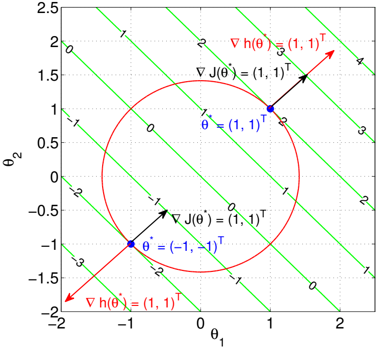
- Feasibility domain (four inequality constraints) :

$$\Omega(\boldsymbol{\theta}) = \left\{\boldsymbol{\theta} \in \mathbb{R}^2\,;\ -4 \leq \boldsymbol{\theta}_1 \leq -1 \text{ and } -3 \leq \boldsymbol{\theta}_2 \leq 4\right\}$$

Example 2

Example

$$\min_{\theta \in \mathbb{R}^2} \quad \theta_1 + \theta_2$$
$$\text{s.c.} \quad \theta_1^2 + \theta_2^2 - 2 = 0$$



- An equality constraint
- Domain of feasibility: a circle with center at $\mathbf{0}$ and diameter equals to 2
- The optimal solution is obtained for $\boldsymbol{\theta}^* = \begin{pmatrix} -1 & -1 \end{pmatrix}^\top$ and we have $J(\boldsymbol{\theta}^*) = -2$

Optimality

- How to assess a solution of the primal problem?
- Do we have optimality conditions similar to those of unconstrained optimization?

Notion of Lagrangian

Primal problem $\mathcal{P}$

$$
\begin{array}{lll}
\min_{\boldsymbol{\theta}\in\mathbb{R}^d} & J(\boldsymbol{\theta}) & \\
 & f_i(\boldsymbol{\theta}) = 0 \quad \forall i = 1, \cdots, n & n \text{ equality constraints} \\
\text{s.c.} & g_j(\boldsymbol{\theta}) \leq 0 \quad \forall j = 1, \cdots, m & m \text{ inequality constraints}
\end{array}
$$

$\boldsymbol{\theta}$ is called primal variable

Principle of Lagrangian

- Each constraint is associated to a scalar parameter called Lagrange multiplier
- Equality constraint $f_i(\boldsymbol{\theta}) = 0$: we associate $\lambda_i \in \mathbb{R}$
- Inequality constraint $g_j(\boldsymbol{\theta}) \leq 0$: we associate $\mu_j \geq 0$[a]
- Lagrangian allows to transform the problem with constraints into a problem without constraints with additional variables: λ_i and μ_j.

[a] Beware of the sense of the inequality $g_j(\boldsymbol{\theta}) \leq 0$

Lagrangian formulation

$$\begin{array}{lll} \min_{\boldsymbol{\theta}\in\mathbb{R}^d} & J(\boldsymbol{\theta}) & \\ & f_i(\boldsymbol{\theta}) = 0 & \forall i = 1,\cdots,n \\ \text{s.c.} & g_j(\boldsymbol{\theta}) \leq 0 & \forall j = 1,\cdots,m \end{array}$$

Associated Lagrange parameters

None

λ_i any real number $\forall i = 1,\cdots,n$

$\mu_j \geq 0 \quad \forall j = 1,\cdots,m$

Lagrangien

The Lagrangian is defined by :

$$\mathcal{L}(\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\mu}) = J(\boldsymbol{\theta}) + \sum_{i=1}^{n}\lambda_i f_i(\boldsymbol{\theta}) + \sum_{j=1}^{m}\mu_j g_j(\boldsymbol{\theta}) \quad \text{avec} \quad \mu_j \geq 0, \forall j = 1,\cdots,m$$

- Parameters $\lambda_i, i = 1,\cdots,n$ and $\mu_j, j = 1,\cdots,m$ are called dual variables
- Dual variables are unknown parameters to be determined

Examples

Example 1

$$\begin{array}{lll} \min\limits_{\theta \in \mathbb{R}^2} & \theta_1 + \theta_2 & \\ \text{s.c.} & \theta_1^2 + 2\theta_2^2 - 2 \leq 0 & \text{inequality constraint} \\ \text{s.c.} & \theta_2 \geq 0 & \text{inequality constraint (mind the type of inequality)} \end{array}$$

Lagrangian : $\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\theta}) = (\theta_1 + \theta_2) + \mu_1(\theta_1^2 + 2\theta_2^2 - 2) + \mu_2(-\theta_2), \quad \mu_1 \geq 0, \mu_2 \geq 0$

Example 2

$$\begin{array}{lll} \min\limits_{\boldsymbol{\theta} \in \mathbb{R}^3} & \frac{1}{2}\left(\theta_1^2 + \theta_2^2 + \theta_3^2\right) & \\ \text{s.c.} & \theta_1 + \theta_2 + 2\theta_3 = 1 & \text{equality constraint} \\ & \theta_1 + 4\theta_2 + 2\theta_3 = 3 & \text{equality constraint} \end{array}$$

Lagrangian :

$$\mathcal{L}(\lambda, \boldsymbol{\theta}) = \frac{1}{2}\left(\theta_1^2 + \theta_2^2 + \theta_3^2\right) + \lambda_1(\theta_1 + \theta_2 + 2\theta_3 - 1) + \lambda_2(\theta_1 + 4\theta_2 + 2\theta_3 - 3), \quad \lambda_1, \lambda_2 \text{ any}$$

Necessary optimality conditions

Assume that J, f_i, g_j are differentiable functions. Let $\boldsymbol{\theta}^*$ be a feasible solution to the problem $\mathcal{P}$. Then there exists dual variables $\lambda_i^*, i = 1, \cdots, n$, $\mu_j^*, j = 1, \cdots, m$ such that the following conditions are satisfied.

Karush-Kuhn-Tucker (KKT) Conditions

Stationarity	$\nabla \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\theta}) = 0 \quad ie$ $\nabla J(\boldsymbol{\theta}) + \sum_{i=1}^n \lambda_i \nabla f_i(\boldsymbol{\theta}) + \sum_{j=1}^m \mu_j \nabla g_j(\boldsymbol{\theta}) = 0$
Primal feasibility	$f_i(\boldsymbol{\theta}) = 0 \qquad \forall i = 1, \cdots, n$ $g_j(\boldsymbol{\theta}) \leq 0 \qquad \forall j = 1, \cdots, m$
Dual feasibility	$\mu_j \geq 0 \qquad \forall j = 1, \cdots, m$
Complementary slackness	$\mu_j g_j(\boldsymbol{\theta}) = 0 \qquad \forall j = 1, \cdots, m$

Example

$$\begin{aligned} \min_{\theta \in \mathbb{R}^2} \quad & \frac{1}{2}(\theta_1^2 + \theta_2^2) \\ \text{s.c.} \quad & \theta_1 - 2\theta_2 + 2 \leq 0 \end{aligned}$$

- Lagrangian : $\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \frac{1}{2}(\theta_1^2 + \theta_2^2) + \mu(\theta_1 - 2\theta_2 + 2), \quad \mu \geq 0$
- KKT Conditions
  - Stationarity: $\nabla_\theta \mathcal{L}(\mu, \boldsymbol{\theta}) = 0 \quad \Rightarrow \quad \begin{cases} \theta_1 = -\mu \\ \theta_2 = -2\mu \end{cases}$
  - Primal feasibility : $\theta_1 - 2\theta_2 + 2 \leq 0$
  - Dual feasibility : $\mu \geq 0$
  - Complementary slackness : $\mu(\theta_1 - 2\theta_2 + 2) = 0$
- Remarks on the complementary slackness
  - If $\theta_1 - 2\theta_2 + 2 < 0$ (inactive constraint) $\Rightarrow \mu = 0$ (no penalty required as the constraint is satisfied)
  - If $\mu > 0 \Rightarrow \theta_1 - 2\theta_2 + 2 = 0$ (active constraint)

Duality

Dual function

Let $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ be the lagrangian of the primal problem $\mathcal{P}$ with $\mu_j \geq 0$. The dual function corresponding to it is $\mathcal{D}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu})$

Theorem [Weak duality]

Let $p^* = \min_\theta \{J(\boldsymbol{\theta})$ such that $f_i(\boldsymbol{\theta}) = 0 \; \forall i$ and $g_j(\boldsymbol{\theta}) \leq 0 \, \forall j\}$ be the optimum value (supposed finite) of the problem $\mathcal{P}$. Then, for any value of $\mu_j \geq 0, \forall j$ and $\lambda_i, \forall i$, we have

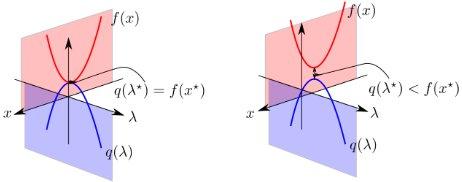
$$\mathcal{D}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^*$$

Dual problem

- The weak duality indicates that the dual function $\mathcal{D}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_\theta \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is a lower bound of p^*
- Bridge the gap: maximize the dual w.r.t. dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ to make this lower bound close to p^*

Dual problem

$$\max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad \mathcal{D}(\boldsymbol{\lambda}, \boldsymbol{\mu})$$
$$\text{s.c.} \quad \lambda_j \geq 0 \qquad \forall j = 1, \cdots, m$$



strong duality   weak duality

http://www.onmyphd.com/?p=duality.theory

Interest of the dual problem

Remarks

- Transform the primal problem into an equivalent dual problem possibly much simpler to solve
- Solving the dual problem can lead to the solution of the primal problem
- Solving the dual problem gives the optimal values of the Lagrange multipliers

Example : inequality constraints

$$\begin{array}{ll} \min\limits_{\theta \in \mathbb{R}^2} & \frac{1}{2}(\theta_1^2 + \theta_2^2) \\ \text{s.c.} & \theta_1 - 2\theta_2 + 2 \leq 0 \end{array}$$

- Lagrangian : $\mathcal{L}(\boldsymbol{\theta}, \mu) = \frac{1}{2}(\theta_1^2 + \theta_2^2) + \mu(\theta_1 - 2\theta_2 + 2), \quad \mu \geq 0$
- Stationarity of the KKT Condition :

$$\nabla_\theta \mathcal{L}(\mu, \boldsymbol{\theta}) = 0 \qquad \Rightarrow \qquad \begin{cases} \theta_1 = -\mu \\ \theta_2 = 2\mu \end{cases} \qquad (1)$$

- Dual function $\mathcal{D}(\mu) = \min_\theta L(\boldsymbol{\theta}, \mu)$: by substituting (1) in $\mathcal{L}$ we obtain

$$\mathcal{D}(\mu) = -\frac{5}{2}\mu^2 + 2\mu$$

- Dual problem : $\max_\mu \mathcal{D}(\mu) \quad \text{s.c.} \quad \mu \geq 0$
- Dual solution

$$\nabla \mathcal{D}(\mu) = 0 \Rightarrow \mu = \frac{2}{5} \quad \text{(that satisfies } \mu \geq 0\text{)} \qquad (2)$$

- Primal solution : (2) and (1) lead to $\boldsymbol{\theta} = \begin{pmatrix} -\frac{2}{5} & \frac{4}{5} \end{pmatrix}^\top$

Constrained optimization with equality constraints

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^3} \quad \frac{1}{2}\left(\theta_1^2+\theta_2^2+\theta_3^2\right)$$
$$\text{s.c.} \quad \theta_1+\theta_2+2\theta_3-1=0$$
$$\theta_1+4\theta_2+2\theta_3-3=0$$

Matrix form

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^3} \quad \frac{1}{2}\boldsymbol{\theta}^\top\boldsymbol{\theta} \quad \text{s.c.} \quad \mathbf{A}\boldsymbol{\theta}-\mathbf{b}=0$$

with $\mathbf{A}=\begin{pmatrix}1 & 1 & 2\\ 1 & 4 & 2\end{pmatrix}$, $\mathbf{b}=\begin{pmatrix}1\\3\end{pmatrix}$

- Lagrangian : $\mathcal{L}(\boldsymbol{\lambda},\boldsymbol{\theta})=\frac{1}{2}\boldsymbol{\theta}^\top\boldsymbol{\theta}+\boldsymbol{\lambda}^\top(\mathbf{A}\boldsymbol{\theta}-\mathbf{b})\,,\quad \boldsymbol{\lambda}\in\mathbb{R}^2$
- KKT Conditions
  - Stationarity : $\nabla_\theta\mathcal{L}(\boldsymbol{\lambda},\boldsymbol{\theta})=0 \quad\Rightarrow\quad \boldsymbol{\theta}=-\mathbf{A}^\top\boldsymbol{\lambda} \quad (1)$
- Dual function : by substituting (1) in $\mathcal{L}$ we get

$$\mathcal{D}(\boldsymbol{\lambda})=-\frac{1}{2}\boldsymbol{\lambda}^\top\mathbf{A}\mathbf{A}^\top\boldsymbol{\lambda}-\lambda^\top\mathbf{b}$$

- Dual Problem : $\max_{\boldsymbol{\lambda}}\mathcal{D}(\boldsymbol{\lambda})$
- Dual Solution

$$\nabla\mathcal{D}(\boldsymbol{\lambda})=0\Rightarrow\boldsymbol{\lambda}=-\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\mathbf{b} \qquad (3)$$

- Primal Solution : (3) in (1) gives $\boldsymbol{\theta}=\mathbf{A}^\top\left(\mathbf{A}\mathbf{A}^\top\right)^{-1}\mathbf{b}$

Convex constrained optimization

$$
\begin{array}{lll}
\min_{\boldsymbol{\theta}\in\mathbb{R}^d} & J(\boldsymbol{\theta}) & \\
 & f_i(\boldsymbol{\theta}) = 0 & \forall i = 1, \cdots, n \\
\text{s.c.} & g_j(\boldsymbol{\theta}) \leq 0 & \forall j = 1, \cdots, m
\end{array}
$$

Convexity condition

- J is a convex function
- f_i are linear $\forall i = 1, n$
- g_j are convex functions $\forall j = 1, m$

Problems of interest

- Linear Programming (LP)
- Quadratic Programming (QP)
- Off-the-shelves toolboxes exist for those problems (Gurobi, Mosek, CVX ...)

QP convex problem

Standard form

$$\begin{array}{llll} \min\limits_{\theta \in \mathbb{R}^d} & \frac{1}{2}\boldsymbol{\theta}^\top \mathbf{G}\boldsymbol{\theta} + \mathbf{q}^\top \boldsymbol{\theta} + r & & \\ \text{s.c.} & \mathbf{a}_i^\top \boldsymbol{\theta} = b_i & \forall i = 1, \cdots, n & \text{affine equality constraint} \\ & \mathbf{c}_j^\top \boldsymbol{\theta} \geq d_j & \forall j = 1, \cdots, m & \text{linear inequality constraints} \end{array}$$

with $\mathbf{q}, \mathbf{a}_i, \mathbf{c}_j \in \mathbb{R}^d$, $\mathbf{d}_i$ and d_j real scalar values and $\mathbf{G} \in \mathbb{R}^{d \times d}$ a **positive definite matrix**

Examples

SVM Problem

$$\begin{array}{ll} \min\limits_{\theta \in \mathbb{R}^2} & \dfrac{1}{2}(\theta_1^2 + \theta_2^2) \\ \text{s.c.} & \theta_1 - 2\theta_2 + 2 \leq 0 \end{array}$$

$$\begin{array}{lll} \min_{\theta, b\mathbb{R}} & \frac{1}{2}\|\boldsymbol{\theta}\|^2 & \\ \text{s.t.} & y_i(\boldsymbol{\theta}^\top x_i + b) \geq 1 & \forall i = 1, N \end{array}$$

Summary

- Optimization under constraints: involved problem in the general case
- Lagrangian: allows to reduce to an unconstrained problem via Lagrange multipliers
- Lagrange multipliers
  - to a constraint corresponds a multiplier $\Rightarrow$
  - act as a penalty if the corresponding constraints are violated
- Optimally: Stationary condition + feasibility conditions + Complementary conditions
- Duality: provides lower bound on the primal problem. Dual problem sometimes easier to solve than primal.

A reference book