Introduction to constrained optimization

Gilles Gasso

INSA Rouen - ASI Departement
Laboratory LITIS

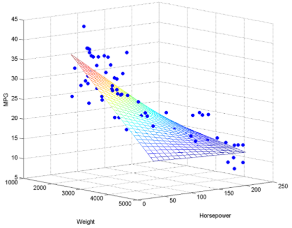
October 7, 2024

Plan

Constrained optimization problems

- Where do these problems come from?
- How to formulate them?

Example 1: sparse Regression



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

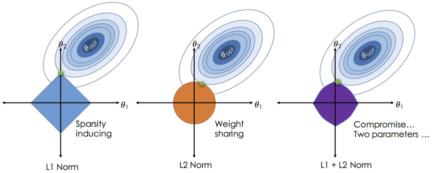
Determination of a sparse $\boldsymbol{\theta}$

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

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

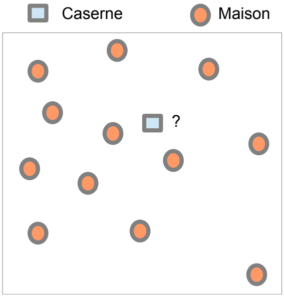
$$\text{s.t. } \|\boldsymbol{\theta}\|_p \leq k$$

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 $z_i = [x_i \ y_i]^\top$
- Let θ be the coordinates of the firehouse
- Minimize the distance from the firehouse to the farthest house

Problem formulation

$$\min_{\theta} \max_{i=1,\cdots,n} \|\theta - z_i\|^2$$

Equivalent problem

$$\min_{t\in\mathbb{R},\theta\in\mathbb{R}^2} t$$
$$\text{s.t. } \|\theta - z_i\|^2 \leq t \quad \forall i = 1, \cdots, n$$

Constrained optimization problem

Notations and assumptions

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

Primal problem $\mathcal{P}$

$\min_{\boldsymbol{\theta} \in \mathbb{R}^d}$	$J(\boldsymbol{\theta})$		objective function	
s.t.	$f_i(\boldsymbol{\theta}) = 0$	$\forall i = 1, \cdots, n$	n	Equality Constraints
	$g_j(\boldsymbol{\theta}) \leq 0$	$\forall j = 1, \cdots, m$	m	Inequality Constraints

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

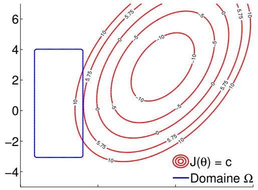
- $\boldsymbol{\theta}_0$ *is feasible if* $\boldsymbol{\theta}_0 \in \text{dom}J$ *and* $\boldsymbol{\theta}_0 \in \Omega(\boldsymbol{\theta})$ *ie* θ_0 *fulfills all the constraints and* $J(\theta_0)$ *has a finite value*
- $\boldsymbol{\theta}^*$ *is a global solution of the problem if* $\boldsymbol{\theta}^*$ *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

$$\min_{\boldsymbol{\theta}} \quad 0.9\theta_1^2 - 0.74\theta_1\theta_2 + 0.75\theta_1^2 - 5.4\theta_1 - 1.2\theta_2$$

$$\text{s.t.} \quad -4 \le \theta_1 \le -1$$

$$-3 \le \theta_2 \le 4$$



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

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

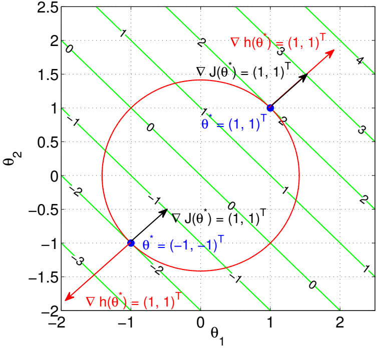
- Feasibility domain (four inequality constraints):

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

Example 2

Example

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



- An equality constraint
- Domain of feasibility: a circle with center at 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 $\mathcal{P}$

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

$\boldsymbol{\theta}$ 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.

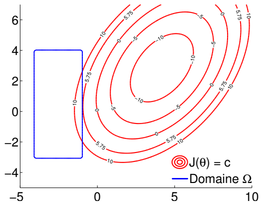
[a] Beware of the type of inequality i.e. $g_j(\boldsymbol{\theta}) \leq 0$

Example

$$\min_{\theta \in \mathbb{R}^2} \quad 0.9\theta_1^2 - 0.74\theta_1\theta_2 + 0.75\theta_1^2 - 5.4\theta_1 - 1.2\theta_2$$

$$\text{s.t.} \quad -4 \leq \theta_1 \leq -1$$

$$-3 \leq \theta_2 \leq 4$$



Constraints (inequality)

1. $-4 \leq \theta_1 \Leftrightarrow -\theta_1 - 4 \leq 0$
2. $\theta_1 \leq -1 \Leftrightarrow \theta_1 + 1 \leq 0$
3. $-3 \leq \theta_2 \Leftrightarrow -\theta_2 - 3 \leq 0$
4. $\theta_2 \leq 4 \Leftrightarrow -\theta_2 - 4 \leq 0$

Related Lagrange Parameters

1. $\mu_1 \geq 0$
2. $\mu_2 \geq 0$
3. $\mu_3 \geq 0$
4. $\mu_4 \geq 0$

Lagrangian formulation

		Associated Lagrange parameters
$\min_{\boldsymbol{\theta}\in\mathbb{R}^d}$	$J(\boldsymbol{\theta})$	None
	$f_i(\boldsymbol{\theta}) = 0 \quad \forall i = 1, \cdots, n$	λ_i any real number $\quad \forall i = 1, \cdots, n$
s.c.	$g_j(\boldsymbol{\theta}) \leq 0 \quad \forall j = 1, \cdots, m$	$\mu_j \geq 0 \qquad \forall j = 1, \cdots, m$

Lagrangien

The Lagrangian is defined by :

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

- **Lagrange 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

$$\begin{aligned}\min_{\theta\in\mathbb{R}^2} \quad & 0.9\theta_1^2 - 0.74\theta_1\theta_2 + 0.75\theta_1^2 - 5.4\theta_1 - 1.2\theta_2 \\ \text{s.t.} \quad & -4 \le \theta_1 \le -1 \quad \text{and} \quad -3 \le \theta_2 \le 4\end{aligned}$$

Lagrangian

$$\begin{aligned}\mathcal{L}(\boldsymbol{\mu},\boldsymbol{\theta}) &= 0.9\theta_1^2 - 0.74\theta_1\theta_2 + 0.75\theta_1^2 - 5.4\theta_1 - 1.2\theta_2 \\ &+ \mu_1(-\theta_1 - 4) + \mu_2(\theta_1 + 1) + \mu_3(-\theta_2 - 3) + \mu_4(-\theta_2 - 4)\end{aligned}$$

with $\mu_1 \ge 0, \mu_2 \ge 0, \mu_3 \ge 0, \mu_4 \ge 0$ (because of inequality constraints)

Example 2

$$\begin{aligned}\min_{\boldsymbol{\theta}\in\mathbb{R}^3} \quad & \frac{1}{2}\left(\theta_1^2 + \theta_2^2 + \theta_3^2\right) \\ \text{s.t.} \quad & \theta_1 + \theta_2 + 2\theta_3 = 1 \quad \text{equality constraint} \\ & \theta_1 + 4\theta_2 + 2\theta_3 = 3 \quad \text{equality constraint}\end{aligned}$$

Lagrangian

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

with $\lambda_1, \lambda_2 \in \mathbb{R}$ (equality constraints)

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 $\quad \nabla\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\theta}) = 0 \quad ie$

$$\nabla J(\boldsymbol{\theta}) + \sum_{i=1}^{n} \lambda_i \nabla f_i(\boldsymbol{\theta}) + \sum_{j=1}^{m} \mu_j \nabla g_j(\boldsymbol{\theta}) = 0$$

Primal feasibility $\quad f_i(\boldsymbol{\theta}) = 0 \qquad \forall i = 1, \cdots, n$

$\quad g_j(\boldsymbol{\theta}) \leq 0 \qquad \forall j = 1, \cdots, m$

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

Complementary slackness $\quad \mu_j g_j(\boldsymbol{\theta}) = 0 \qquad \forall j = 1, \cdots, m$

Example

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

- 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: $\nabla_\theta \mathcal{L}(\mu, \boldsymbol{\theta}) = 0 \quad \Rightarrow \quad \left\{ \begin{array}{l} \theta_1 = -\mu \\ \theta_2 = -2\mu \end{array} \right.$
  - 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 corresponding dual function is defined as*

$$\mathcal{D}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Theorem [Weak duality]

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

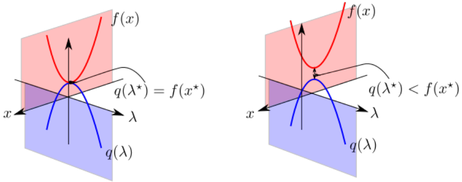
$$\mathcal{D}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^*$$

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

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



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

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

- 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 :

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

- 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

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

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

Convex constrained optimization

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

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 ...)

Gurobi Optimization   mosek   CVX RESEARCH

QP convex problem

Standard form

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

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

Examples

SVM Problem

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

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

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