Constrained optimization

Pierre Hellier

Formulation

General problem formulation for constrained optimization

Solve for $x^* = arg\min_{\mathbf{x}\in\mathbb{R}^n} L(x)$
with $h_j(x) = 0, \forall j = 1, \ldots, p$
and $g_i(x) < 0, \forall i = 1, \ldots, q$
where h_j and g_i define equality and inequality constraints

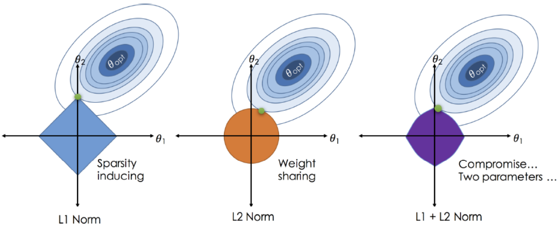
Introduction

- Where do these problems come from?
- What is the connection with machine learning?

Example 1: Sparse regression

Problem formulation:

- Predict a scalar output y out of multiple observations $x \in \mathbb{R}^n$
- Linear model $f(x) = \theta^T x$, where $\theta \in \mathbb{R}^n$ are the parameters of the model
- Enforce sparsity (too many measurements were conducted, blind observations): $||\theta||_p \leq k$ where $||\theta||_p = \sum_i |\theta_i|^p$



Example 2: Resource allocation

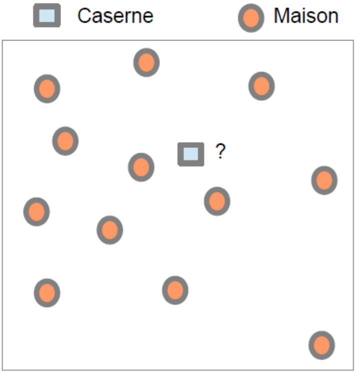
Problem formulation:

- Observed distribution of *source* and *targets*
- Minimize distance between distributions
- Primal problem:

$$arg \min_{\theta} \max_{i} ||\theta - z_i||^2$$

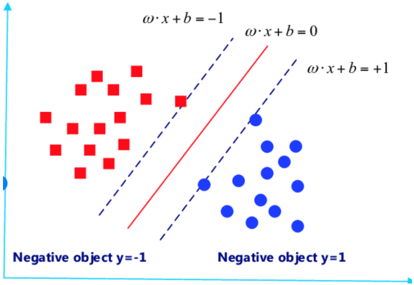
- Dual problem:

$$\min_{t\in\mathbb{R},\theta\in\mathbb{R}^2} t, s.t. \forall i, ||\theta - z_i||^2 \leq t$$



Example 3: Support vector machines (1)

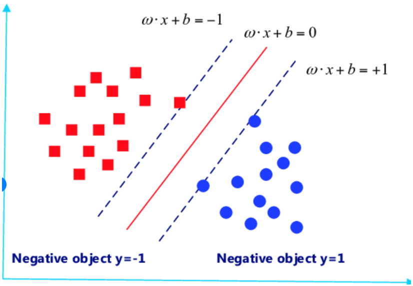
- Hyperplan as linear classifier, $\forall i, x_i \in \mathbb{R}^n, y_i \in \{-1, 1\}$
- $\hat{y}_i = sign(f(x_i))$ with $f(x) = w^t x + b$
- Ambiguity in possible solutions
- Frontier is the locus of points x where $f(x) = 0$



Example 3: Support vector machines (2)

Maximize the margin

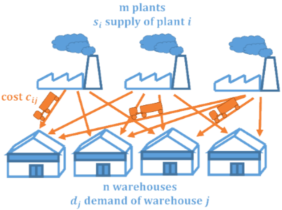
- $\forall x, \frac{|w^t x+b|}{||w||}$ is the distance to the decision frontier
- Maximal margin $\delta = \frac{2}{||w||}$
- Maximize the margin leads to minimize $||w||^2$
- SVM problem
  - $arg\min_w ||w||^2$
  - under the constraint $\forall i, y_i f(x_i) \geq 1$



Example 4: Optimal transport

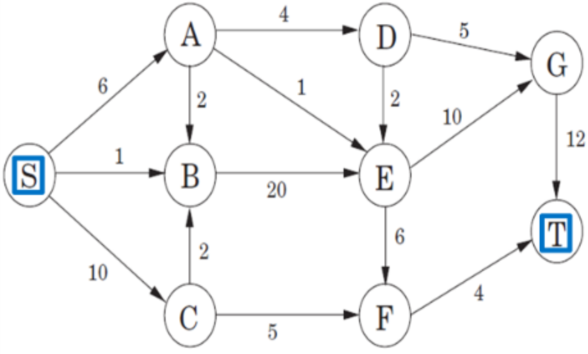
- m factories produce s_i quantities of goods
- n stores sell d_j amount of goods
- Cost C_{ij} of transporting goods from factory i to store j
- Optimal transport: find to optimal way to transport goods
- $\min_X \sum_{i,j} C_{ij} X_{ij}$ such that $\sum_i X_{ij} \geq d_j$ (fulfill the needs of store), $\sum_j X_{ij} \leq s_i$ (do not exceed supply of each factory) and $X_{ij} \geq 0$

A general resource allocation problem, proposed by Kantorovich in 1942 (Nobel prize economy 1975).



Example 5: Max flow problem

- Maximize flow from source S to target T (can be water, oil, *etc.*)
- Edges have a limited capacity
- Nodes cannot store fluid $\Leftrightarrow$ flow input $=$ flow output



Dual problem and KKT

Feasibility conditions

- The problem might not admit a solution!
- Feasibility domain

$$\Omega(x) = \{x \in \mathbb{R}^n | h_j(x) = 0, \forall j, g_i(x) < 0, \forall i\}$$

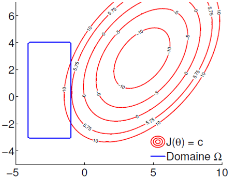
Example (1)

Problem

$$\arg\min_{\theta_1,\theta_2} 0.9\theta_1^2 + 0.75\theta_2^2 - 0.74\theta_1\theta_2 - 5.4\theta_1 - 1.2\theta_2$$

such that

$$-4 \leq \theta_1 \leq -1 \text{ and } -3 \leq \theta_2 \leq 4$$



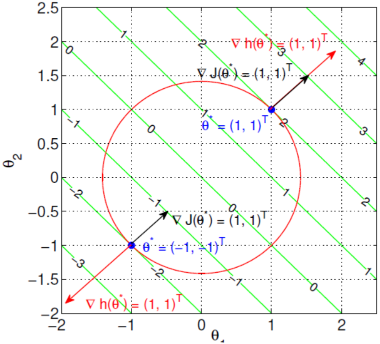
Example (2)

Problem

$$\arg\min_{\theta_1,\theta_2} \theta_1 + \theta_2$$

such that

$$\theta_1^2 + \theta_2^2 - 2 = 0$$



Dual lagrangian problem

Idea: transform the problem with constraints into a problem without constraints and additional variables. We define the lagrangian problem as the minimization of:

$$\mathcal{L}(x, u, v) = F(x) + \sum_{i=1}^{i=q} u_i g_i(x) + \sum_{j=1}^{j=p} v_j h_j(x)$$

Where

- u and v are vectors called Lagrange multipliers of dual variables
- $u_i \geq 0, \forall i$ (positivity constraint)
- $D(u, v) = \inf_x \mathcal{L}(x, u, v)$

$D, \mathcal{L}$ are lower bounds $D(u, v) \leq \mathcal{L}(x, u, v) \leq F(x)$

Exercice (1)

Write the Lagrangian formulation of the following problem:

Problem

$$\mathit{arg}\min_{\theta_1,\theta_2} \theta_1 + \theta_2$$

such that

$$\theta_1^2 + 2\theta_2^2 - 2 \leq 0$$

and

$$\theta_2 \geq 0$$

Solution

$$\mathcal{L}(x, \mu_1, \mu_2) = (\theta_1 + \theta_2) + \mu_1(\theta_1^2 + 2\theta_2^2 - 2) + \mu_2(-\theta_2)$$

with $\mu_1 \geq 0$, $\mu_2 \geq 0$

Exercice (2)

Write the Lagrangian formulation of the following problem:

Problem

$$\mathit{arg}\min_{\theta \in \mathbb{R}^3} \frac{1}{2}(\theta_1^2 + \theta_2^2 + \theta_3^2)$$

such that

$$\theta_1 + \theta_2 + 2\theta_3 = 1$$

and

$$\theta_1 + 4\theta_2 + 2\theta_3 = 3$$

Solution

$$\mathcal{L}(x, u, v) = \frac{1}{2}(\theta_1^2+\theta_2^2+\theta_3^2)+\lambda_1(\theta_1+\theta_2+2\theta_3-1)+\lambda_2(\theta_1+4\theta_2+2\theta_3-3)$$

without constraints on λ_i

Duality gap

Recall, minimization of:

$$\mathcal{L}(x, u, v) = F(x) + \sum_{i=1}^{i=q} u_i g_i(x) + \sum_{j=1}^{j=p} v_j h_j(x)$$

Where

- u and v are vectors called Lagrange multipliers of dual variables
- $u_i \geq 0, \forall i$ (positivity constraint)
- $D(u, v) = \inf_x \mathcal{L}(x, u, v)$

$D, \mathcal{L}$ are lower bounds $D(u, v) \leq \mathcal{L}(x, u, v) \leq F(x)$

Duality gap

We define the duality gap as:

$$F(x) - D(u, v) \geq 0$$

, where $D(u, v) = \inf_x \mathcal{L}(x, u, v)$

- If the duality gap is zero for a triplet x^*, u^*, v^*, then x^* is optimal for the primal and u^*, v^* are optimal for the dual
- If $F^* = D^*$, the problem is said to have strong duality
- **Stater's constraint qualification**: if the primal problem is convex and there exists a feasible solution (meeting the equality and inequality constraints), then strong duality holds

Karush-Kuhn-Tucker (KKT) conditions

Important conditions at equilibrium x^*, u^*, v^*:

1. $\nabla \mathcal{L}(x^*, u^*, v^*) = 0$ (stationarity condition)
2. $g_i(x^*) \leq 0, h_j(x^*) = 0$ (primal feasibility condition)
3. $u_i^* \geq 0$ (dual feasibility condition)
4. $\forall i, u_i^* g_i(x^*) = 0$ (complementary slackness condition)

KKT theorem

The KKT conditions are necessary and sufficient to find optimal solutions of primal and dual problems.

Practicalities

How to find a solution (if it exists)

- Write the Lagragian
- Write the KKT conditions
- Compute analytic solution x^* as function of u and v
- Express the dual problem and solve if easier
- Use KKT to recover solution x^*

Exercice

Problem

Solve

$$\min_{\theta} \frac{1}{2}(\theta_1^2 + \theta_2^2)$$

such that

$$\theta_1 - 2\theta_2 + 2 \leq 0$$

Solution

- Lagrangian $\mathcal{L}(\theta, \mu) = \frac{1}{2}(\theta_1^2 + \theta_2^2) + \mu(\theta_1 - 2\theta_2 + 2)$ with $\mu \geq 0$
- Stationarity $\nabla_\theta \mathcal{L} = 0 \Rightarrow \theta_1 = -\mu, \theta_2 = 2\mu$
- Dual function $D(\mu) = -\frac{5}{2}\mu^2 + 2\mu$
- Dual problem, $max_{\mu, \mu \geq 0} D \Rightarrow \mu = \frac{2}{5}$
- Back to primal, $\theta_1 = -\frac{2}{5}$ and $\theta_2 = \frac{4}{5}$

Linear projection and PCA as a constrained optimization problem

- PCA is a linear projection operator. Input data $X = [\mathbf{x}_1, \ldots, \mathbf{x}_m] \in \mathbb{R}^{nm}$. Output (or latent) data $Y = [\mathbf{y}_1, \ldots, \mathbf{y}_m] \in \mathbb{R}^{dm}$ with $d < n$
- Objectives: data visualization, dimension reduction
- Projection: $Y = Q^t X$, where $Q^t \in \mathbb{R}^{dn}$
- Reconstruction $\hat{X} = QY$, where $Q \in \mathbb{R}^{nd}$
- Reconstruction can be used for generative models (out of scope)

Exercice

Write and solve the lagrangian associated to PCA on dimension $d = 1$

Problem

Tips:

- Write the 1D projection of the vectors onto one direction u, assuming data is centered (zero mean)
- Write the maximization of the variance along the direction, as a function of the input data covariance
- Write the lagragian of the problem, under the constraint of the unit norm for u
- Compute the derivative of the lagrangian w.r.t. u
- Solve for the lagrangian multiplier knowing that the variance is maximal

$1d$ projection

- Suppose first we want to project data along a line given by direction u. Goal: maximize variance of latent data to maximize information.
- Variance σ_u

$$\sigma_u = \frac{1}{m}\sum_i (\mathrm{u}^t \mathrm{x_i})^2$$

$$\sigma_u = \mathrm{u}^t \left(\frac{1}{m}\sum_i \mathrm{x_i}\mathrm{x_i}^t \right) \mathrm{u}$$

$$\sigma_u = \mathrm{u}^t \Sigma_X \mathrm{u}$$

- Maximize variance under constraint of unit vector: $\mathrm{u}^t\mathrm{u} = 1$

$1d$ projection

- Lagrange multiplier $L(\mathrm{u},\alpha) = \mathrm{u}^t\Sigma_X\mathrm{u} - \alpha(\mathrm{u}^t\mathrm{u} - 1)$
- $\frac{\partial L}{\partial \mathrm{u}} = \Sigma_X\mathrm{u} - \alpha\mathrm{u}$
- a solution verifies $\Sigma_X\mathrm{u} = \alpha\mathrm{u}$
- u is an eigenvector of Σ_X associated with eigenvalue α
- $\mathrm{u}^t\Sigma_X\mathrm{u} = \alpha\mathrm{u}^t\mathrm{u} = \alpha$
- Hence, solution is the eigenvector associated with largest eigenvalue

PCA generalization

- PCA amount to a diagonalization of the covariance matrix
- New basis formed by axes of largest variance
- $\Sigma_X = U \Lambda U^t = \sum_a \lambda_a \mathrm{u_a} \mathrm{u_a}^t$
- Projection in dimension d: $\lambda_1, \ldots, \lambda_d$ are the d largest eigenvalues associated with eigenvectors $\mathrm{u_1}, \ldots, \mathrm{u_d}$
- $U_d = [\mathrm{u_1}, \ldots, \mathrm{u_d}] \in \mathbb{R}^{nd}$, $\forall i, \mathrm{y_i} = U_d^t \mathrm{x_i}$
- Projected data is decorrelated (data whitening):

$$\frac{1}{m} \sum_i \mathrm{y_i} \mathrm{y_i}^t = \frac{1}{m} \sum_i U_d^t \mathrm{x_i} \mathrm{x_i}^t U_d = U_d^t \Sigma_X U_d = U_d^t U \Lambda U^t U_d = \Lambda$$

Linear equality constraints

Special case where we only have linear equality constraints, i.e.:

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

with $A \in \mathbb{R}^{n \times p}$ defining p linearly independent constraints

- Eliminate constraints with linear algebra, $x = x_{ker} + y$, where $x_{ker} \in Ker(A)$
- $\{x|Ax = b\} = \{kz + \hat{x}|z \in \mathbb{R}^{n-p}\}$ where $Im(k) = Ker(A)$ and $\hat{x}$ such that $A\hat{x} = b$
- In practice: use `scipy.linalg.null_space` to compute null space
- Equivalent unconstrained problem:

$$z^* = \min_z f(kz + \hat{x})$$

and $x^* = kz^* + \hat{x}$

Interior point method

From analytical solution to numerical solution

Problem

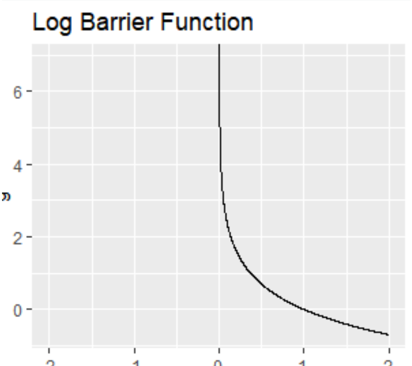
What happens is analytical solution cannot be found?

Numerical solving!

Log-barrier function

Proposed by Renegar (1988) and Gonzaga (1989). Main idea: replace inequality $g_i \leq 0$ with log-barrier function

$$min_x f(x) + \frac{1}{\delta} \sum_i -\log(-g_i(x))$$

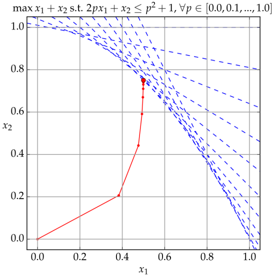


Interior point solver

Main idea: iterative solutions that stay in the admissible domain

Algorithm:

- Init with feasible x and $\delta > 0; \mu > 1$
  1. $x \to x(\delta)$ (needs a solver for this smooth problem, see unconstrained part of the course)
  2. $\delta = \mu\delta$
- until convergence, where $\lim_{\delta \to \infty} = x^*$
- Mostly polynomial time



(Credit By Gjacquenot - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=39949878)

Linear and quadratic programming

Linear Programming (LP) formulation

$$x^* = arg \max_x c^T x$$

such that $Ax \leq b$ and $x \geq 0$

- Problems can be expressed with positive variables with reformulation tricks
- Interpretation in economy: A factory produces n goods produced (x is the vector of quantities), where the selling prize vector is c. Each good needs some basic material for production (matrix A), and the stock of material is b. Objective: maximize revenue.

History of LP problems

- 1700: Fourier proposed an inefficient *Fourier-Motzkin* elimination method
- 1930: Kantorovic resource allocation problem
- 1947: Von Neuman the duality problem
- 1947: Dantzig the simplex algorithm
- 1979: Khachiyan the ellipsoid method (in practice, slow)
- 1984: Karmarkar the interior point method

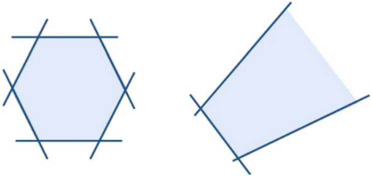
Solvability

There are different cases:

- The problem is unfeasible because the constraints do not intersect
- The problem is unbounded, so as the solution
- There is an infinite number of optimal solutions
- There is a unique optimal solution

Geometrical view

- Hyperplan: $\{x | a^T x = b\}$
- Half plane: $\{x | a^T x \leq b\}$
- Polyhedron: intersection of a finite set of half planes
- Polytope: a bounded polyhedron
- Extreme, or vertex: a point that cannot be expressed as a linear combination of two points in polytope
- A vertex is always on the boundary (the reverse is not true)



Fundamental theorem of LP

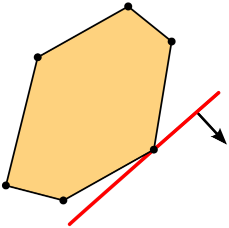
If LP has a unique optimal solution, then the solution is a vertex of the polytope
Demonstration:

- If x^* is the unique optimal solution and not a vertex
- $\exists u \in \mathbb{R}^n, \epsilon \in \mathbb{R}$ such that $x_1 = x^* + \epsilon u$ and $x_2 = x^* - \epsilon u$ are in the polytope
- Necessarily, either $c^T x_1$ or $c^T x_2$ are greater than $c^T x^*$, and x^* is not optimal

Solutions

Properties of the solution:

- Solution x^* is always on one side of the polytope
- Convex problem, but could have infinity of solutions (one side of the polytope)
- if $A \in \mathbb{R}^{p \times n}$, there are at most p components of x^* that are non-zero
- Interesting if $p < n$, for instance for the resource allocation (n products with p material), there is an optimal plan using at most 3 products



How to solve the problem?

In practice

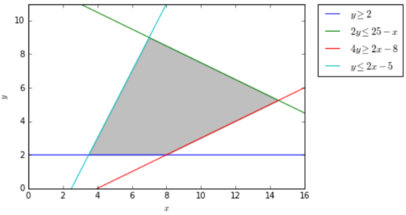
- Try analytic solving (works only for small simple problems)
- Express an analytic solution with dual formulation
- Numerical solving with simplex or interior point

Exercice

Problem

Solve the following problem: max $z = 4x + 3y$ subject to $x \geq 0$, $y \geq 2$, $2y \leq 25 - x$,$4y \geq 2x - 8$,$y \leq 2x - 5$
Tips: Draw admissible domain, inspect vertices

Solution



Compute value at four vertices

- $V_1, x = 8, y = 2, Z = 38$
- $V_2, x = 7, y = 9, Z = 55$
- $V_3, x = 14.5, y = 5.25, Z = 73.75$
- $V_4, x = 3.5, y = 2, Z = 20$

Solution only possible for small size problems...

Dual LP formulation

$$x^* = arg \max_x c^T x$$

such that $Ax \leq b$ and $x \geq 0$

- Instead of producing goods, sell material to another factory (offering price per-item $y \geq 0$)
- The offer can only be accepted if $A^T y \geq c$ (otherwise the first factory should produce goods out of material)
- The second factory aims at minimizing the cost $b^T y$

Dual LP:

$$\min_y b^T y$$

such that $A^T y \geq c$ and $y \geq 0$.

Dual gap

- The dual problem is an upper bound of the primal that we aim at minimizing
- The dual problem may (or may not) be easier to solve
- Construct the dual: each primal constrint becomes a dual variable, and signs are reversed

Dual LP:

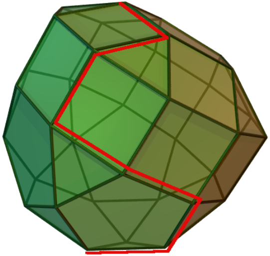
$$\min_y b^T y$$

such that $A^T y \geq c$ and $y \geq 0$.

Simplex solver

Intuition:

- Init x^* with some admissible point
- Iteratively move along the polytope edges to find better solution
- Invented by Dantzig around 1947

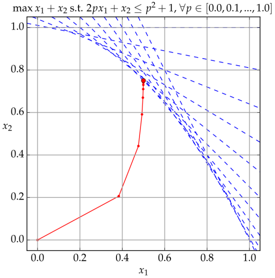


Interior point solver

solve

$$\max_x \delta c^T x - \sum_i \log(a_i^T x - b_i)$$

- Classical solver `scipy.optimize.linprog` and library `cvxopt`
- Simplex optimizes on the border of the polytope, IP solves within it
- Polynomial complexity (better than simplex in theory)



Practical session

Linear programming notebook