Modern Optimization Techniques

Lucas Rego Drumond

Information Systems and Machine Learning Lab (ISMLL)
Institute of Computer Science
University of Hildesheim, Germany

Duality

Outline

1. Constrained Optimization
2. Duality
3. KKT Conditions

Outline

1. Constrained Optimization

2. Duality

3. KKT Conditions

Constrained Optimization Problems

A **constrained optimization problem** has the form:

$$\begin{array}{rll} \text{minimize} & f_0(\mathbf{x}) & \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, & i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 0, & i = 1, \ldots, p \end{array}$$

Where:

- $f_0, \ldots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$
- $h_1, \ldots, h_p : \mathbb{R}^n \rightarrow \mathbb{R}$
- An optimal $\mathbf{x}^*$

Convex Constrained Optimization Problems

A constrained optimization problem:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_j(\mathbf{x}) = 0, \quad i = 1, \ldots, p \end{aligned}$$

is convex iff:

- $f_0, \ldots, f_m$ are convex
- $h_1, \ldots, h_p$ are affine

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & H\mathbf{x} = \mathbf{b} \end{aligned}$$

Linear Programming

$$
\begin{array}{rl}
\text{minimize} & \mathbf{c}^T\mathbf{x} \\
\text{subject to} & \mathbf{a}_i^T\mathbf{x} \leq b_i \quad i = 1, \ldots, m
\end{array}
$$

- No analytical solution
- There are reliable algorithms available

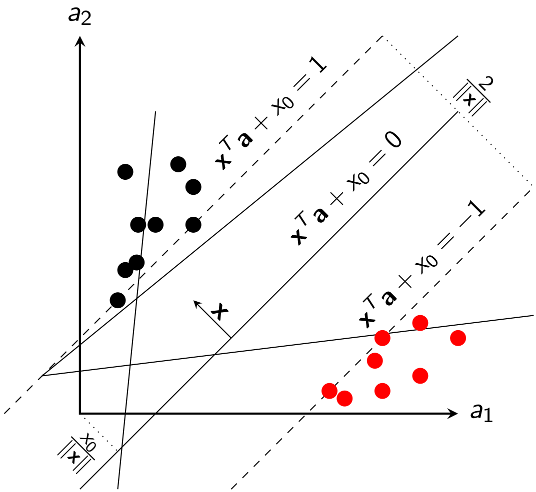
Quadratic Programming

$$\begin{array}{rl} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \mathbf{a}_i^T\mathbf{x} \leq b_i \quad i = 1, \ldots, m \end{array}$$

where:

- $Q \succeq 0$

Maximum Margin Separating Hyperplanes



Support Vector Machines

If the instances are not completely separable, we can allow some of them to be on the wrong side of the decision boundary

The closer the “wrong” points are to the boundary the better (modeled by slack variables ξ_i

$$\begin{array}{rl} \text{minimize} & \frac{1}{2}||\mathbf{x}||^2 + \gamma \sum_{i=1}^{n} \xi_i \\ \text{subject to} & y_i(x_0 + \mathbf{x}^T \mathbf{a_i}) \geq 1 - \xi_i \quad i = 1, \ldots, n \\ & \xi_i \geq 0 \quad i = 1, \ldots, n \end{array}$$

Outline

Lagrangian

Given a constrained optimization problem in the standard form:

$$
\begin{array}{rll}
\text{minimize} & f_0(\mathbf{x}) & \\
\text{subject to} & f_i(\mathbf{x}) \leq 0, & i = 1, \ldots, m \\
 & h_j(\mathbf{x}) = 0, & i = 1, \ldots, p
\end{array}
$$

We can put the objective function and the constraints in the same expression:

$$f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{p} \nu_j h_j(\mathbf{x})$$

The expression above is not the same original problem. It is called the primal **Lagrangian** of the problem

Lagrangian

The **primal Lagrangian** of a constrained optimization problem is a function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$:

$$L(\mathbf{x}, \lambda, \nu) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

where:

- λ_i and ν_j are called Lagrange multipliers
- λ_i is the Lagrange multiplier associated with the constraint $f_i(\mathbf{x}) \leq 0$
- ν_i is the Lagrange multiplier associated with the constraint $h_i(\mathbf{x}) = 0$

Dual Lagrangian

Be $\mathcal{D}$ the domain of the problem, the **dual Lagrangian** of a constrained optimization problem is a function $g : \times\mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$:

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x}, \lambda, \nu) \\
&= \inf_{\mathbf{x}\in\mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right)
\end{aligned}
$$

where g is concave

Interesting fact: for non-negative λ_i, g is a **lower bound** on $f_0(\mathbf{x}^*)$, i.e:

If $\lambda \succeq 0$, then $g(\lambda, \nu) \leq f_0(\mathbf{x}^*)$

Dual Lagrangian

Proof of the lower bound property of:

$$\begin{aligned} g(\lambda, \nu) &= \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, \nu) \\ &= \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right) \end{aligned}$$

for a feasible $\mathbf{x}'$ we have:

- $h_i(\mathbf{x}') = 0$
- $f_i(\mathbf{x}') \leq 0$

thus, with $\lambda \succeq 0$:

$$f_0(\mathbf{x}') \geq L(\mathbf{x}', \lambda, \nu) \geq \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible $\mathbf{x}'$ we have $f_0(\mathbf{x}^*) \geq g(\lambda, \nu)$

Computing the dual

$$\begin{aligned} \text{minimize} \quad & (\mathbf{y} - A\mathbf{x})^2 \\ \text{subject to} \quad & -\mathbf{x}_i \leq 0 \\ & H\mathbf{x} = \mathbf{b} \end{aligned}$$

Lagrangian: $L(\mathbf{x}, \lambda, \nu) = (\mathbf{y} - A\mathbf{x})^2 + \sum_{i=1}^{n} -\lambda_i \mathbf{x}_i + \nu(H\mathbf{x} - \mathbf{b})$
Dual Lagrangian: Minimize L over $\mathbf{x}$

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda\nu) = -2A^T(\mathbf{y} - A\mathbf{x}) - \lambda + H^T\nu = 0$$
$$\mathbf{x} = \frac{1}{2}(H^T\nu - \lambda - 2A^T\mathbf{y})A^TA^{-1}$$

Substitute in L to get g

Least-norm solution of linear equations

$$\begin{aligned} \text{minimize} \quad & \mathbf{x}^T\mathbf{x} \\ \text{subject to} \quad & H\mathbf{x} = \mathbf{b} \end{aligned}$$

Lagrangian: $L(\mathbf{x}, \nu) = \mathbf{x}^T\mathbf{x} + \nu(H\mathbf{x} - \mathbf{b})$
Dual Lagrangian: Minimize L over $\mathbf{x}$

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \nu) = 2\mathbf{x} + H^T\nu = 0$$
$$\mathbf{x} = -\frac{1}{2}H^T\nu$$

Substituting in L to get g: $g(\nu) = -\frac{1}{4}\nu^T HH^T\nu - b^T\nu$

The dual problem

Once we know how to compute the dual, we are interested in computing the *best* lower bound on $f_0(\mathbf{x}^*)$:

$$\begin{array}{rl} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

where:

- this is a convex optimization problem (g is concave)
- d^* is the optimal value of g

Weak and Strong Duality

Say p^* is the optimal value of f_0 and d^* is the optimal value of g

We have **weak duality** when: $d^* \leq p^*$

- Always holds
- Can be useful to find informative lower bounds for difficult problems

We have **strong duality** when: $d^* = p^*$

- Does not always hold
- Holds for a range of convex problems
- Properties that guarantee strong duality are called constraint qualifications

Slater's Condition

If the following primal problem

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \qquad i = 1, \ldots, m \\ & H\mathbf{x} = \mathbf{b} \end{aligned}$$

is:

- convex
- strictly feasible, i.e.

$$\exists \mathbf{x} : f_i(\mathbf{x}) < 0 \quad i = 1, \ldots, m, \qquad H\mathbf{x} = \mathbf{b}$$

then strong duality holds for this problem

Duality Gap

How close is the value of the dual lagrangian to the primal objective?

Given a primal feasible $\mathbf{x}$ and a dual feasible λ, ν, the **duality gap** is given by:

$$f_0(\mathbf{x}) - g(\lambda, \nu)$$

Since $g(\lambda, \nu)$ is a lower bound on f_0:

$$f_0(\mathbf{x}) - f_0(\mathbf{x}^*) \leq f_0(\mathbf{x}) - g(\lambda, \nu)$$

If the duality gap is zero, then $\mathbf{x}$ is primal optimal

This is a useful stopping criterion since if $f_0(\mathbf{x}) - g(\lambda, \nu) \leq \epsilon$, then we are sure that $f_0(\mathbf{x}) - f_0(\mathbf{x}^*) \leq \epsilon$

Outline

Complementary Slackness

Assume strong duality where $\mathbf{x}^*$ is the primal optimal and (λ^*, ν^*) is dual optimal:

$$\begin{aligned} f_0(\mathbf{x}^*) = g(\lambda^*, \nu^*) &= \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\ &\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \\ &\leq f_0(\mathbf{x}^*) \end{aligned}$$

hence:

$$f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) = f_0(\mathbf{x}^*)$$

and $\mathbf{x}^*$ minimizes $L(\mathbf{x}, \lambda^*, \nu^*)$

Complementary Slackness

Assume we have a problem with strong duality where $\mathbf{x}^*$ is the primal optimal and (λ^*, ν^*) is dual optimal.

$$f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) = f_0(\mathbf{x}^*)$$

From this we can derive the **complementary slackness**:

For $i = 1, \ldots, m$

$$\lambda_i^* f_i(\mathbf{x}^*) = 0$$

Which means that

- If $\lambda_i^* > 0$ then $f_i(\mathbf{x}^*) = 0$
- If $f_i(\mathbf{x}^*) < 0$ then $\lambda_i = 0$

Karush-Kuhn-Tucker (KKT) Conditions

The following conditions are called the KKT conditions:

1. Primal feasibility: $f_i(\mathbf{x}) \leq 0$ and $h_j(\mathbf{x}) = 0$ for all i, j
2. Dual feasibility: $\lambda \succeq 0$
3. Complementary Slackness: $\lambda_i f_i(\mathbf{x}) = 0$ for all i
4. Stationarity: $\nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i \nabla h_i(\mathbf{x}) = 0$

If strong duality holds and $\mathbf{x}, \lambda, \nu$ are optimal, then they **must** satisfy the KKT conditions

If $\mathbf{x}, \lambda, \nu$ satisfy the KKT conditions, then $\mathbf{x}$ is the primal solution and (λ, ν) is the dual solution