Modern Optimization Techniques

3. Equality Constrained Optimization / 3.1. Duality

Lars Schmidt-Thieme

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

Syllabus

Outline

1. Constrained Optimization
2. Duality
3. Karush-Kuhn-Tucker Conditions

Outline

Constrained Optimization Problems

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

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_p(\mathbf{x}) = 0, \quad p = 1, \ldots, P \\ & h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \end{aligned}$$

where:

- $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is called the **objective** or **cost function**,
- $g_1, \ldots, h_P : \mathbb{R}^N \rightarrow \mathbb{R}$ are called **equality constraints**,
- $h_1, \ldots, h_Q : \mathbb{R}^N \rightarrow \mathbb{R}$ are called **inequality constraints**,
- a feasible, optimal $\mathbf{x}^*$ exists

Constrained Optimization Problems

A **convex constrained optimization problem**:

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_p(\mathbf{x}) = 0, \quad p = 1, \ldots, P \\ & h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \end{aligned}$$

is **convex** iff:

- f, the **objective function** is convex,
- $g_1, \ldots, g_P$ the **equality constraint functions** are affine: $g_p(x) = \mathbf{a}_p^T \mathbf{x} - b_p$, and
- $h_1, \ldots, h_Q$ the **inequality constraint functions** are convex.

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{a}_p^T \mathbf{x} - b_p = 0, \quad p = 1, \ldots, P \\ & h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \end{aligned}$$

Linear Programming

A convex problem with an

- **affine objective** and
- **affine constraint** functions

is called **Linear Program (LP)**.

Standard form LP:

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{a}_p^T \mathbf{x} = b_p, \quad p = 1, \ldots, P \\ & \mathbf{x} \geq 0 \end{aligned}$$

Inequality form LP:

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{a}_q^T \mathbf{x} \leq b_q, \quad q = 1, \ldots, Q \end{aligned}$$

- No analytical solution
- There are specialized algorithms available

Quadratic Programming

A convex problem with

- a **quadratic objective** and
- **affine constraint** functions

is called **Quadratic Program (QP)**.

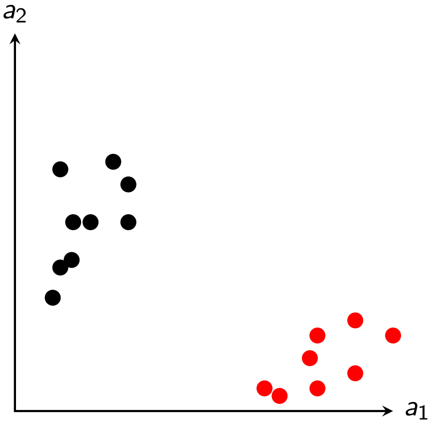
Inequality form LP:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^T C\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{a}_q^T\mathbf{x} \leq b_q, \quad q = 1, \ldots, Q \end{aligned}$$

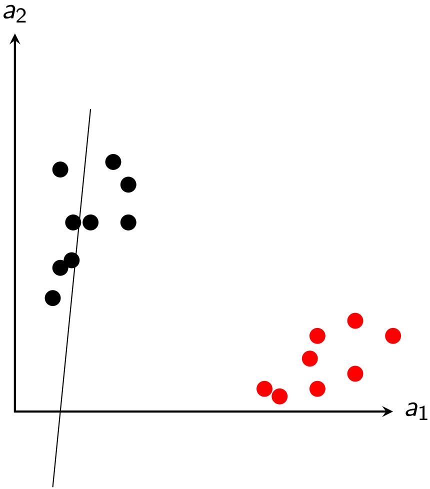
where:

- $C \succ 0$ pos.def.,
- $C = 0$, a special case: linear programs.

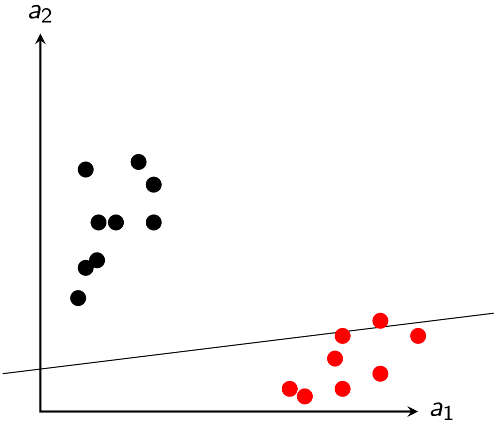
Example: Maximum Margin Separating Hyperplanes



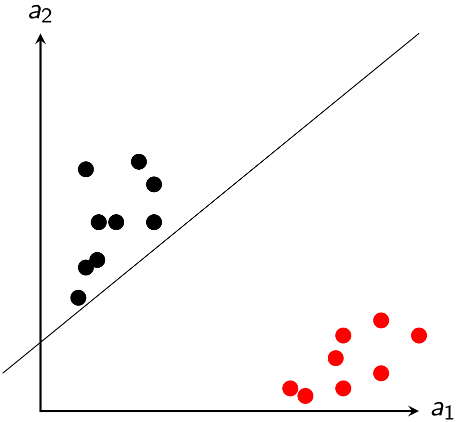
Example: Maximum Margin Separating Hyperplanes



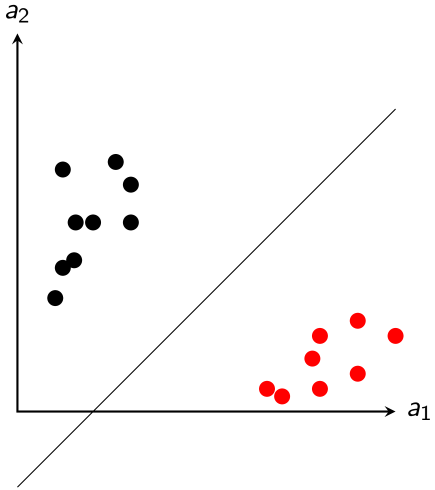
Example: Maximum Margin Separating Hyperplanes



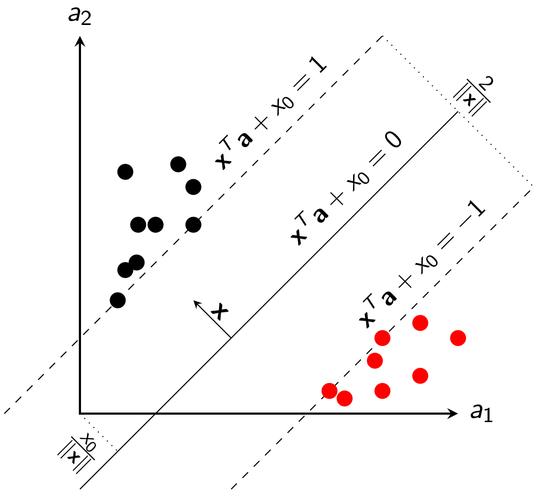
Example: Maximum Margin Separating Hyperplanes



Example: Maximum Margin Separating Hyperplanes



Example: Maximum Margin Separating Hyperplanes



Example: 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).

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}||\mathbf{x}||^2 + \gamma \sum_{i=1}^{l} \xi_i \\ \text{subject to} \quad & y_i(a_0 + \mathbf{x}^T \mathbf{a_i}) \geq 1 - \xi_i \quad i = 1, \dots, l \\ & \xi_i \geq 0 \quad i = 1, \dots, l \end{aligned}$$

Outline

1. Constrained Optimization
2. Duality
3. Karush-Kuhn-Tucker Conditions

Lagrangian

Given a constrained optimization problem in the standard form:

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_p(\mathbf{x}) = 0, \quad p = 1, \ldots, P \\ & h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \end{aligned}$$

We can put the objective function and the constraints in a joint function called **primal Lagrangian**:

$$f(\mathbf{x}) + \sum_{p=1}^{P} \nu_p \, g_p(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q \, h_q(\mathbf{x})$$

Primal Lagrangian

The **primal Lagrangian** of a constrained optimization problem is a function $L : \mathbb{R}^N \times \mathbb{R}^P \times \mathbb{R}^Q \to \mathbb{R}$:

$$L(\mathbf{x}, \nu, \lambda) := f(\mathbf{x}) + \sum_{p=1}^{P} \nu_p \, g_p(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q \, h_q(\mathbf{x})$$

where:

- ν_p and λ_q are called **Lagrange multipliers**.
  - ν_p is the Lagrange multiplier associated with the constraint $g_p(\mathbf{x}) = 0$
  - λ_q is the Lagrange multiplier associated with the constraint $h_q(\mathbf{x}) \leq 0$.

Dual Lagrangian

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

$$g(\nu, \lambda) := \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \nu, \lambda)$$
$$= \inf_{\mathbf{x} \in \mathcal{D}} \left(f(\mathbf{x}) + \sum_{p=1}^{P} \nu_p \, g_p(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q \, h_q(\mathbf{x}) \right)$$

- g is concave.
  - as infimum over concave (affine) functions
- for non-negative λ_q, g is a **lower bound** on $f(\mathbf{x}^*)$:

$$g(\nu, \lambda) \leq f(\mathbf{x}^*) \quad \text{for } \lambda \geq 0$$

Dual Lagrangian / Proof

Proof of the lower bound property of:

$$g(\nu, \lambda) := \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \nu, \lambda)$$
$$= \inf_{\mathbf{x} \in \mathcal{D}} \left(f(\mathbf{x}) + \sum_{p=1}^{P} \nu_p \, g_p(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q \, h_q(\mathbf{x}) \right)$$

for any feasible **x** we have:

- $g_p(\mathbf{x}) = 0$
- $h_q(\mathbf{x}) \leq 0$

thus, with $\lambda \geq 0$:

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

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

Least-norm solution of linear equations

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

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

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

  - Substituting $\mathbf{x}$ in L we get g:

$$g(\nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T\nu$$

The dual problem

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

$$\begin{aligned} \text{maximize} \quad & g(\nu, \lambda) \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned}$$

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
and d^* is the optimal value of g

Weak duality: $d^* \leq p^*$

- always holds
- can be useful to find informative lower bounds for difficult problems

Strong duality: $d^* = p^*$

- does not always hold
- but holds for a range of convex problems
- properties that guarantee strong duality are called **constraint qualifications**

Slater's Condition

If the following primal problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b} \\ & h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \end{aligned}$$

is:

- convex and
- **strictly feasible**, i.e.

$$\exists \mathbf{x} : \quad A\mathbf{x} = \mathbf{b} \quad \text{and} \quad h_q(\mathbf{x}) < 0, \quad q = 1, \ldots, Q$$

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 ν, λ,
the **duality gap** is given by:

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

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

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

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

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

Outline

1. Constrained Optimization
2. Duality
3. Karush-Kuhn-Tucker Conditions

Complementary Slackness

Assume strong duality:

- $\mathbf{x}^*$ is primal optimal and
- (ν^*, λ^*) is dual optimal.

$$f(\mathbf{x}^*) = g(\nu^*, \lambda^*) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f(\mathbf{x}) + \sum_{p=1}^{P} \nu_p^* g_p(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q^* h_q(\mathbf{x}) \right)$$

$$\leq f(\mathbf{x}^*) + \sum_{p=1}^{P} \nu_p^* g_p(\mathbf{x}^*) + \sum_{q=1}^{Q} \lambda_q^* h_q(\mathbf{x}^*)$$

$$\leq f(\mathbf{x}^*)$$

hence

$$f(\mathbf{x}^*) + \sum_{p=1}^{P} \nu_p^* g_p(\mathbf{x}^*) + \sum_{q=1}^{Q} \lambda_q^* h_q(\mathbf{x}^*) = f(\mathbf{x}^*)$$

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

Complementary Slackness

Assume strong duality:

- $\mathbf{x}^*$ is primal optimal and
- (ν^*, λ^*) is dual optimal.

$$f(\mathbf{x}^*) + \sum_{p=1}^{P} \nu_p^* g_p(\mathbf{x}^*) + \sum_{q=1}^{Q} \lambda_q^* h_q(\mathbf{x}^*) = f(\mathbf{x}^*)$$

$\rightsquigarrow$ **complementary slackness**:

$$\lambda_q^* h_q(\mathbf{x}^*) = 0, \quad q = 1, \ldots, Q$$

which means that

- If $\lambda_q^* > 0$, then $h_q(\mathbf{x}^*) = 0$
- If $h_q(\mathbf{x}^*) < 0$, then $\lambda_q = 0$

Karush-Kuhn-Tucker (KKT) Conditions

The following conditions on $\mathbf{x}, \nu, \lambda$ are called the KKT conditions:

1. **primal feasibility:** $g_p(\mathbf{x}) = 0$ and $h_q(\mathbf{x}) \leq 0, \quad \forall p, q$
2. **dual feasibility:** $\lambda \geq 0$
3. **complementary slackness:** $\lambda_q \, h_q(\mathbf{x}) = 0, \quad \forall q$
4. **stationarity:** $\nabla f(\mathbf{x}) \quad + \sum_{p=1}^{P} \nu_p \nabla g_p(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q \nabla h_q(\mathbf{x}) = 0$

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