Optimization Techniques for Machine Learning

AMLZC326 · #10 Duality II

Anshid Aboobacker

MOTIVATION

- CH09 showed that weak duality always holds: $d^* \leq p^*$. When does the gap close?
- The answer — **convexity** and **Slater's condition** — unlocks strong duality and the KKT conditions
- KKT conditions power SVMs, LASSO, and neural network regularisation — three of the most important ML optimisation problems

Learning Objectives

By the end of this lecture you should be able to:

- Define convex sets and convex functions; recognise key sufficient conditions for convexity
- State Slater's condition and explain when strong duality holds ($d^* = p^*$)
- State and verify the four KKT conditions (stationarity, primal feasibility, dual feasibility, complementary slackness)
- Apply KKT conditions to solve a simple constrained optimisation problem by hand

Table of Contents

From Primal to Dual

$$\underbrace{\min_{x} \max_{\lambda \geq 0} L(x, \lambda)}_{\text{Primal}}$$

vs

$$\underbrace{\max_{\lambda \geq 0} \min_{x} L(x, \lambda)}_{\text{Dual}}$$

Are they equal?

Min-Max Inequality

$$\boxed{\max_y \min_x f(x, y) \leq \min_x \max_y f(x, y)}$$

Left Side

$$\max_y \min_x f(x, y)$$

Minimizer moves first.

Maximizer reacts later.

Right Side

$$\min_x \max_y f(x, y)$$

Maximizer moves first.

Minimizer reacts later.

Changing the order changes the game.

Weak Duality

$$\boxed{d^* \leq p^*}$$

Primal Optimum

$$p^* = \min_x \max_{\lambda \geq 0} L(x, \lambda)$$

Primal gives an upper bound.

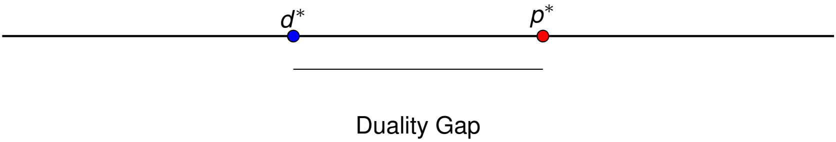
Dual Optimum

$$d^* = \max_{\lambda \geq 0} \min_x L(x, \lambda)$$

Dual gives a lower bound.

Dual can never exceed primal.

Duality Gap



Weak Duality

$$d^* < p^*$$

Gap exists.

Strong Duality

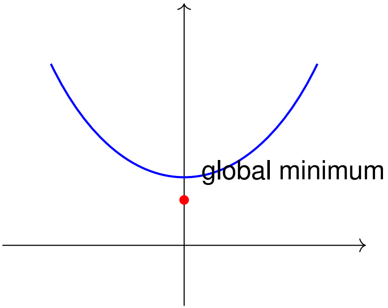
$$d^* = p^*$$

Gap disappears.

Table of Contents

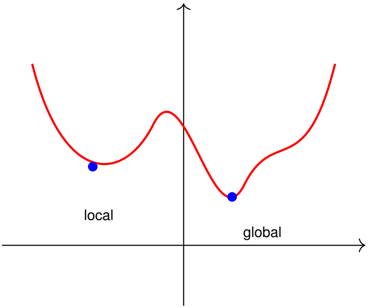
Why Convexity Matters

Convex Optimization



Local minimum = Global minimum

Non-Convex Optimization



Many local minima may exist.

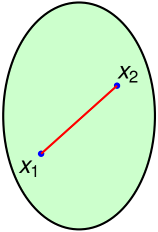
Convexity makes optimization tractable.

CONVEX SETS

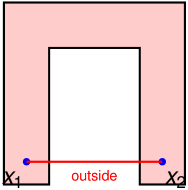
C is convex if for every $x_1, x_2 \in C$ the entire line segment joining them also lies inside C.

$$\lambda x_1 + (1 - \lambda) x_2 \qquad 0 \leq \lambda \leq 1$$

Convex Set



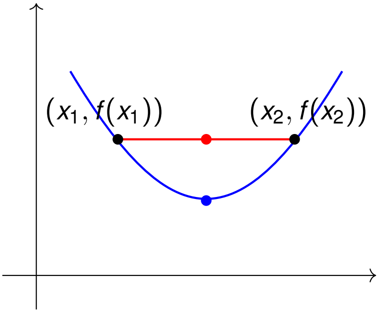
Non-Convex Set



Convex Functions

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

Function value at the midpoint
lies below the straight line joining the endpoints.



Convexity via Derivatives and Hessians

First-Order Characterization:
A differentiable function f is convex if: $f(y) \geq f(x) + \nabla f(x)^T (y - x) \; \forall x, y$.

i.e., the tangent line always lies below the graph.

Single Variable Case

$$f''(x) \geq 0$$

Example:

$$f(x) = x^2$$

$$f''(x) = 2 > 0$$

Hence convex.

Multivariable Case

$$\nabla^2 f(x) \succeq 0$$

Hessian must be positive semidefinite.

$$z^T \nabla^2 f(x) z \geq 0$$

for all vectors z.

Positive curvature $\Longrightarrow$ Convexity

Why Convex Optimization Is Easier

Convex Optimization

- One global minimum
- Gradient descent behaves well
- Easier theoretical guarantees

Non-Convex Optimization

- Many local minima
- Optimization becomes difficult
- Algorithms may get stuck

Machine Learning Context

Deep learning loss landscapes are usually non-convex.

Strong Duality

$$\boxed{d^* = p^*}$$

The dual optimum equals the primal optimum.

Weak Duality

$$d^* \leq p^*$$

A gap may exist.

Strong Duality

$$d^* = p^*$$

Duality gap disappears.

Optimization becomes beautifully symmetric.

Slater's Condition

Suppose:

1. Objective function $f(x)$ is convex
2. Constraint functions $g_i(x)$ are convex
3. There exists a strictly feasible point $\bar{x}$ such that $g_i(\bar{x}) < 0 \qquad \forall i$

Slater's Condition
If these conditions hold, then strong duality holds.

$$d^* = p^*$$

Convexity + Interior Feasible Point $\Longrightarrow$ Zero duality gap

Table of Contents

Why KKT Conditions Matter

$\nabla f(x) = 0$ works only for **unconstrained optimization**

Unconstrained Case

$\min_x f(x)$

Optimal point occurs when:
$\nabla f(x) = 0$

Constrained Case

$\min_x f(x)$ subject to $g_i(x) \leq 0$

Need additional conditions.

KKT conditions generalize $\nabla f(x) = 0$ to constrained optimization.

KKT Conditions

Suppose: x^* is an optimal solution.

1. **Primal Feasibility** $g_i(x^*) \leq 0$
2. **Dual Feasibility** $\lambda_i \geq 0$
3. **Complementary Slackness** $\lambda_i g_i(x^*) = 0$
4. **Stationarity** $\nabla_x L(x^*, \lambda) = 0$

Key Idea

At the optimum, constraints and gradients must balance perfectly.

Complementary Slackness Intuition

$$\boxed{\lambda_i g_i(x^*) = 0}$$

Inactive Constraint

$$g_i(x^*) < 0$$

Constraint is not touching the boundary.

Therefore: $\lambda_i = 0$

No pressure from the constraint.

Active Constraint

$$g_i(x^*) = 0$$

Constraint touches the optimum.

Then: $\lambda_i \geq 0$

Constraint may push against the solution.

$$\text{Inactive constraint} \implies \text{No force}$$

$$\text{Active constraint} \implies \text{Constraint pushes back}$$

A Simple KKT Example

min $f(x) = x^2$ subject to $x \geq 1$

Rewrite constraint: $1 - x \leq 0$

Lagrangian: $L(x, \lambda) = x^2 + \lambda(1 - x)$

Stationarity $\frac{\partial L}{\partial x} = 2x - \lambda = 0$

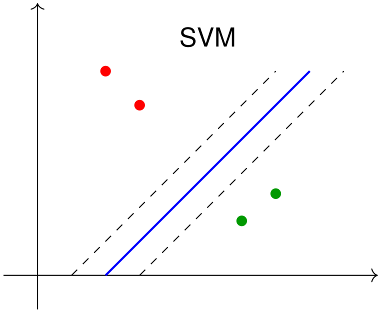
Complementary Slackness $\lambda(1 - x) = 0$

Optimal solution: $x^* = 1, \qquad \lambda^* = 2$

KEY TAKEAWAYS

Applications

- Support Vector Machines
- Regularization
- Sparse optimization
- Robust optimization



Optimization View

Training ML models often means:

$$\min \text{Loss Function}$$

subject to constraints.

Duality is one of the mathematical foundations of modern ML.

Thank you :)