MA289: Mathematics of AI
Lesson XX Outline — XX XXXXXX 202X
United States Military Academy, West Point
Instructor: MAJ Patrick Kuiper

---

1 Lagrange Multipliers and KKT Conditions

- Understand Lagrange Multipliers
- Understand KKT Conditions.

2 Constrained Optimization: Lagrange Multipliers, Duality, and KKT Conditions

2.1 1. Lagrange Multipliers (Equality Constraints)

Consider the constrained optimization problem:

$$\min_{x,y} \ f(x,y) = x^2 + 2y^2 \quad \text{subject to} \quad g(x,y) = x + 2y - 2 = 0$$

Lagrangian:

$$\mathcal{L}(x,y,\lambda) = x^2 + 2y^2 - \lambda(x + 2y - 2)$$

First-order conditions:

$$\frac{\partial \mathcal{L}}{\partial x} = 2x - \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 4y - 2\lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = x + 2y - 2 = 0$$

From the first two equations:

$$x = \frac{\lambda}{2}, \quad y = \frac{\lambda}{2}$$

Substitute into the constraint:

$$\frac{\lambda}{2} + 2\left(\frac{\lambda}{2}\right) = 2 \ \Rightarrow \ \frac{3\lambda}{2} = 2 \ \Rightarrow \ \lambda = \frac{4}{3}$$

Thus:

$$x = y = \frac{2}{3}$$

Interpretation: At the optimum, the gradient of the objective is a linear combination of the constraint gradient.

—

2.2 2. Dual Formulation

Consider:

$$\min_{x} \ f(x) = x^2 + 4x \quad \text{subject to} \quad x \geq 1$$

Rewrite constraint:

$$h(x) = 1 - x \leq 0$$

Lagrangian:

$$\mathcal{L}(x, \lambda) = x^2 + 4x + \lambda(1 - x), \quad \lambda \geq 0$$

Dual function:

$$g(\lambda) = \inf_x \mathcal{L}(x, \lambda)$$

Minimize over x:

$$\frac{d}{dx} = 2x + 4 - \lambda = 0 \;\Rightarrow\; x^*(\lambda) = \frac{\lambda - 4}{2}$$

Substitute back:

$$g(\lambda) = \left(\frac{\lambda - 4}{2}\right)^2 + 4\left(\frac{\lambda - 4}{2}\right) + \lambda\left(1 - \frac{\lambda - 4}{2}\right)$$

Simplifying:

$$g(\lambda) = -\frac{1}{4}\lambda^2 + \lambda - 4$$

Dual problem:

$$\max_{\lambda \geq 0} \; -\frac{1}{4}\lambda^2 + \lambda - 4$$

Taking derivative:

$$-\frac{1}{2}\lambda + 1 = 0 \;\Rightarrow\; \lambda^* = 2$$

Recover primal solution:

$$x^* = \frac{2 - 4}{2} = -1$$

But this violates $x \geq 1$, so the constraint is active:

$$x^* = 1$$

Interpretation: The dual problem optimizes over constraint penalties rather than decision variables.

—

2.3 3. Karush-Kuhn-Tucker (KKT) Conditions

Consider:

$$\min_{x,y} \; f(x, y) = x^2 + y^2 - 2x - 4y \quad \text{subject to} \quad h(x, y) = x + y - 2 \leq 0$$

Lagrangian:

$$\mathcal{L}(x, y, \lambda) = x^2 + y^2 - 2x - 4y + \lambda(x + y - 2)$$

KKT Conditions:

1. **Stationarity:**

$$2x - 2 + \lambda = 0$$

$$2y - 4 + \lambda = 0$$

2. **Primal feasibility:**

$$x + y - 2 \leq 0$$

3. **Dual feasibility:**

$$\lambda \geq 0$$

4. **Complementary slackness:**

$$\lambda(x + y - 2) = 0$$

—

Solve:
From stationarity:

$$x = \frac{2-\lambda}{2}, \quad y = \frac{4-\lambda}{2}$$

Plug into constraint:

$$x + y = \frac{6-2\lambda}{2} = 3 - \lambda$$

Constraint:

$$3 - \lambda - 2 \leq 0 \;\Rightarrow\; 1 - \lambda \leq 0 \;\Rightarrow\; \lambda \geq 1$$

—

Case 1: $\lambda = 0$
Then:

$$x = 1, \quad y = 2$$

Constraint:

$$x + y = 3 > 2 \quad \text{(violates constraint)}$$

Reject.

—

Case 2: Constraint active ($x + y = 2$)
Then:

$$3 - \lambda = 2 \;\Rightarrow\; \lambda = 1$$

Thus:

$$x = \frac{2-1}{2} = \frac{1}{2}, \quad y = \frac{4-1}{2} = \frac{3}{2}$$

—

Final solution:

$$(x^*, y^*) = \left(\frac{1}{2}, \frac{3}{2}\right), \quad \lambda^* = 1$$

—

2.4 4. Summary and Key Insights

- **Lagrange multipliers** transform constrained problems into unconstrained ones by incorporating constraints into the objective.
- **Dual formulation** shifts optimization from decision variables to constraint penalties.
- **KKT conditions** generalize optimality conditions to inequality constraints.
- **Complementary slackness** determines whether constraints are active or inactive.

Conceptual takeaway:

Constraints shape the solution by either binding (active) or disappearing (inactive), and KKT conditions determine which case applies.

2.5 5. Support Vector Machines via Lagrangian and KKT

Consider a simple linearly separable dataset in $\mathbb{R}^2$:

$$x_1 = (2,2),\ y_1 = +1 \quad \text{and} \quad x_2 = (0,0),\ y_2 = -1$$

We solve the hard-margin SVM:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(w \cdot x_i + b) \geq 1$$

—

Step 1: Lagrangian

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{2} \alpha_i \left[y_i(w \cdot x_i + b) - 1\right]$$

—

Step 2: Stationarity (KKT)

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{2} \alpha_i y_i x_i$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^{2} \alpha_i y_i = 0$$

—

Step 3: Solve dual feasibility constraint

$$\alpha_1(1) + \alpha_2(-1) = 0 \Rightarrow \alpha_1 = \alpha_2 = \alpha$$

—

Step 4: Compute w

$$w = \alpha(1)(2,2) + \alpha(-1)(0,0) = (2\alpha, 2\alpha)$$

—

Step 5: Use margin constraints (KKT complementary slackness)
For support vectors:

$$y_i(w \cdot x_i + b) = 1$$

Apply to x_1:

$$(2\alpha, 2\alpha) \cdot (2,2) + b = 1 \Rightarrow 8\alpha + b = 1$$

Apply to x_2:

$$-(b) = 1 \Rightarrow b = -1$$

Substitute:

$$8\alpha - 1 = 1 \Rightarrow \alpha = \frac{1}{4}$$

—

Final solution:

$$w = \left(\frac{1}{2}, \frac{1}{2}\right), \quad b = -1$$

$$f(x) = \frac{1}{2}x_1 + \frac{1}{2}x_2 - 1$$

—

Key KKT Insight:

$$\alpha_i \left[y_i(w \cdot x_i + b) - 1\right] = 0$$

- $\alpha_i > 0 \Rightarrow$ point lies on margin (support vector)
- $\alpha_i = 0 \Rightarrow$ point does not influence w

—

2.6 6. SVM with a Polynomial Kernel

Now consider a dataset that is *not linearly separable* in $\mathbb{R}^2$:

$$x_1 = (1, 0),\ y_1 = +1 \quad x_2 = (0, 1),\ y_2 = +1 \quad x_3 = (0, 0),\ y_3 = -1$$

No linear separator exists.

—

Step 1: Feature mapping

Define:

$$\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$$

Then:

$$\phi(1, 0) = (1, 0, 0), \quad \phi(0, 1) = (0, 1, 0), \quad \phi(0, 0) = (0, 0, 0)$$

Now separable in feature space.

—

Step 2: Kernel representation

Instead of computing $\phi(x)$ explicitly, use the polynomial kernel:

$$K(x_i, x_j) = (x_i \cdot x_j)^2$$

—

Step 3: Dual formulation

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j K(x_i, x_j)$$

subject to:

$$\sum_i \alpha_i y_i = 0, \quad \alpha_i \geq 0$$

—

Step 4: Compute kernel matrix

$$K(x_1, x_1) = 1, \quad K(x_2, x_2) = 1, \quad K(x_3, x_3) = 0$$
$$K(x_1, x_2) = 0, \quad K(x_1, x_3) = 0, \quad K(x_2, x_3) = 0$$

—

Step 5: Solve simplified dual
Constraint:

$$\alpha_1 + \alpha_2 - \alpha_3 = 0$$

Symmetry suggests:

$$\alpha_1 = \alpha_2 = a, \quad \alpha_3 = 2a$$

Plug into dual objective:

$$\max_a \; 2a + 2a - \frac{1}{2}(a^2 + a^2) = 4a - a^2$$

$$\frac{d}{da} = 4 - 2a = 0 \Rightarrow a = 2$$

Thus:

$$\alpha_1 = \alpha_2 = 2, \quad \alpha_3 = 4$$

—

Step 6: Decision function

$$f(x) = \sum_i \alpha_i y_i K(x_i, x) + b$$

$$= 2K(x_1, x) + 2K(x_2, x) - 4K(x_3, x) + b$$

Since $K(x_3, x) = 0$:

$$f(x) = 2(x \cdot (1, 0))^2 + 2(x \cdot (0, 1))^2 + b$$

$$= 2x_1^2 + 2x_2^2 + b$$

—

Interpretation:
The decision boundary is:

$$2x_1^2 + 2x_2^2 + b = 0$$

This is a **nonlinear (circular)** boundary in input space.

—

2.7 7. Key Takeaways

- SVM is solved via **Lagrangian optimization** and **KKT conditions**.
- The **dual formulation** expresses the solution as a weighted sum of training points.
- **Support vectors** arise from complementary slackness.
- **Kernels** replace dot products, enabling nonlinear boundaries without explicit feature mapping.

Conceptual Summary:

Linear separator in feature space $\Rightarrow$ nonlinear separator in input space via kernels.