

# Constrained Optimization

Bailin Deng

Cardiff University



# A 2D Example

- For  $f, g : \mathbb{R}^2 \mapsto \mathbb{R}$ ,  $c \in \mathbb{R}$ ,

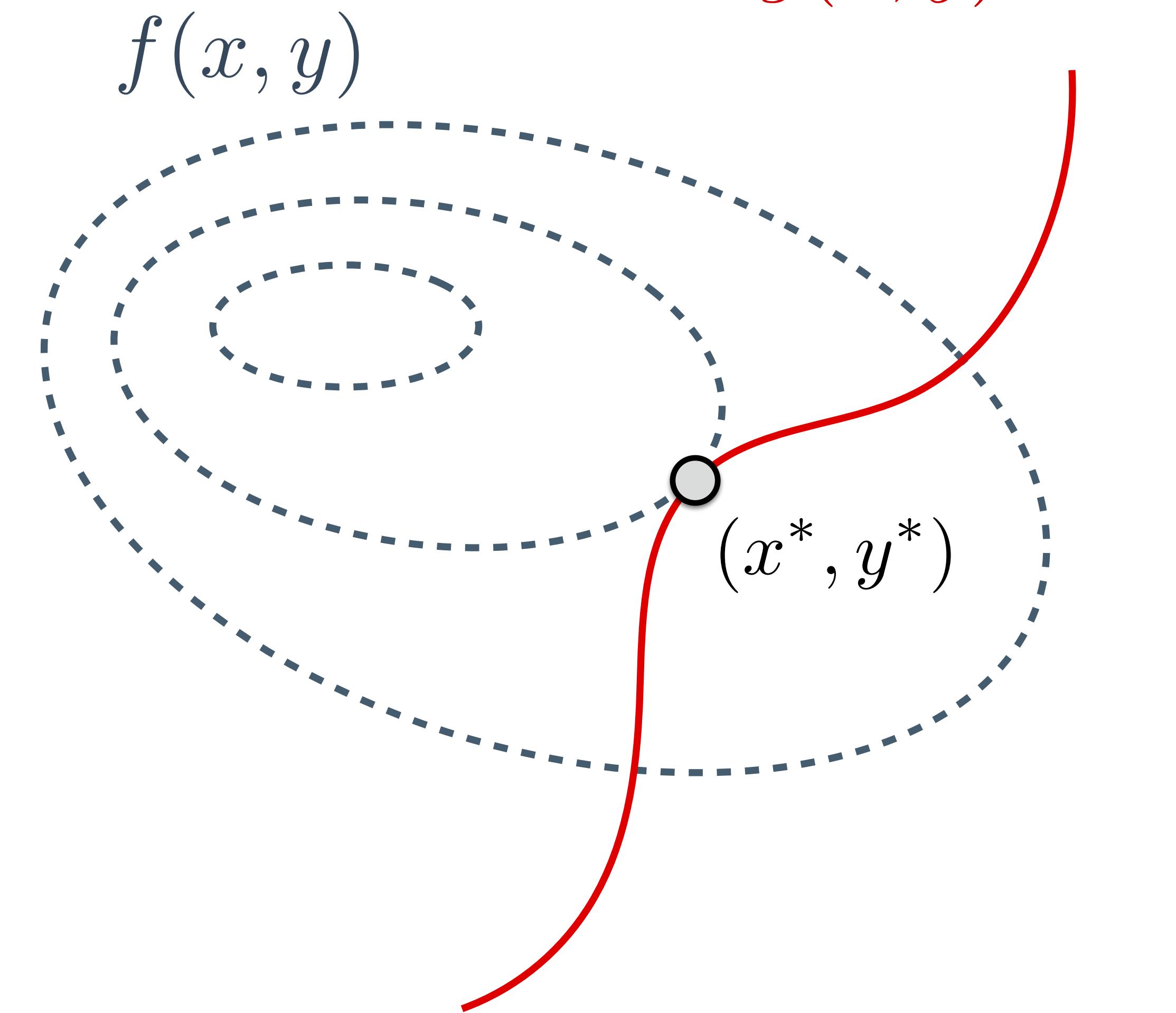
$$\begin{aligned} & \min_{x,y} f(x, y) \\ \text{s.t. } & g(x, y) = 0 \end{aligned}$$

# Optimality Conditions

- Feasibility condition:

$$g(x^*, y^*) = 0$$

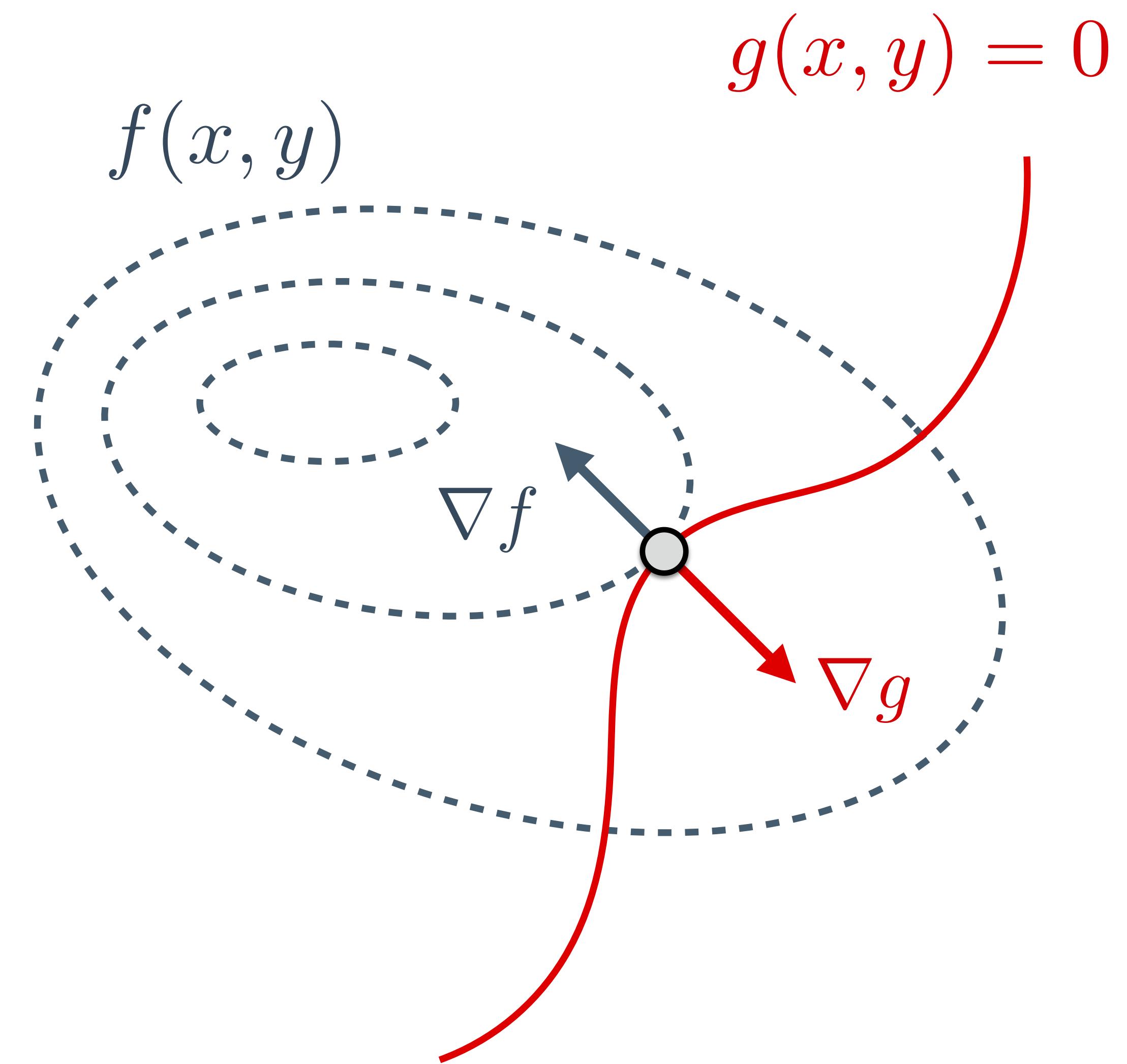
$$g(x, y) = 0$$



# Optimality Conditions

- Necessary condition: gradients  $\nabla f$ ,  $\nabla g$  are linearly dependent

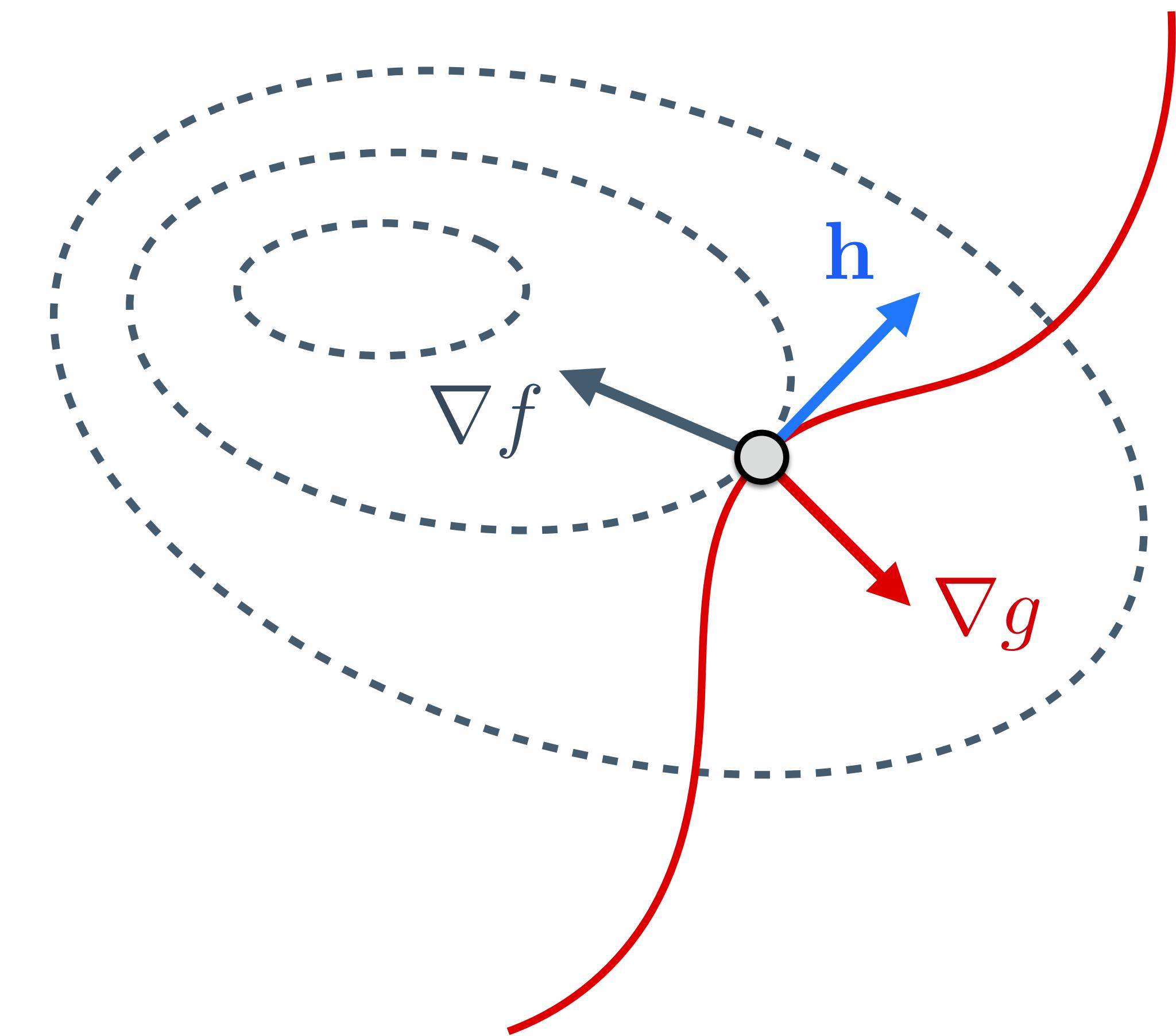
$$\exists \lambda : \nabla f(x^*, y^*) = \lambda \nabla g(x^*, y^*)$$



# Optimality Conditions

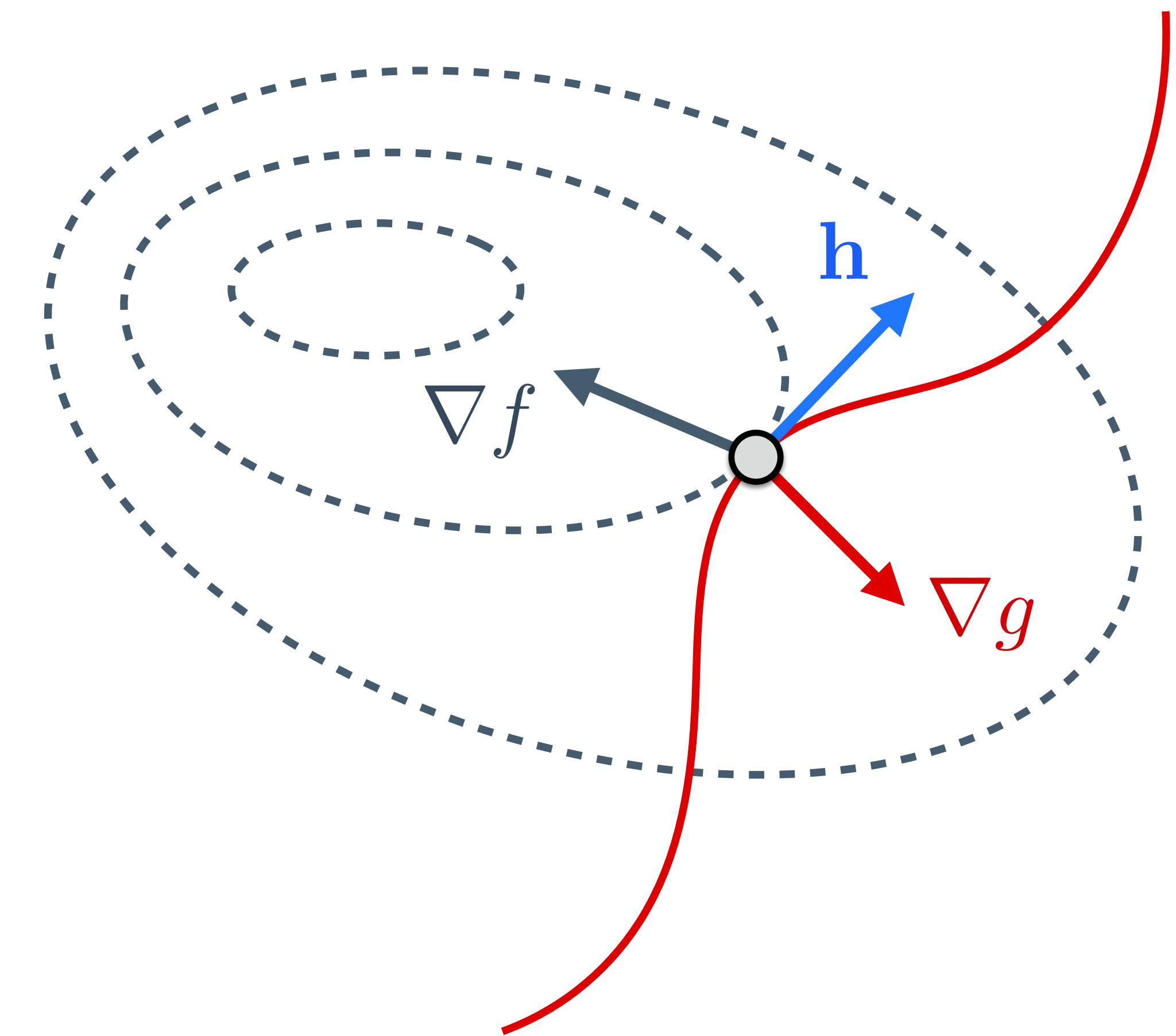
- Proof by contradiction:
  - Suppose  $\nabla f, \nabla g$  linearly independent
  - There exists  $h$  such that

$$\nabla f \cdot h < 0, \quad \nabla g \cdot h = 0$$



# Optimality Conditions

- Proof by contradiction:
  - Suppose  $\nabla f, \nabla g$  linearly independent
  - There exists  $h$  such that
$$\nabla f \cdot h < 0, \quad \nabla g \cdot h = 0$$
  - Infinitesimal change along  $h$  satisfies constraint but decreases  $f$



# High-Dimensional Case

- For  $f, g_1, \dots, g_m : \mathbb{R}^n \mapsto \mathbb{R}$

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

# High-Dimensional Case

- For  $f, g_1, \dots, g_m : \mathbb{R}^n \mapsto \mathbb{R}$

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

- Necessary condition for solution  $\mathbf{x}^*$ : linear dependence of  $\nabla f, \nabla g_1, \dots, \nabla g_m$

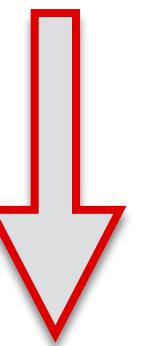
$$\exists \lambda_1, \dots, \lambda_m : \nabla f(\mathbf{x}^*) = \sum_{I=1}^m \lambda_i \nabla g_i(\mathbf{x}^*)$$

# Method of Lagrange Multipliers

- Lagrangian for equality-constrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

$$\text{s.t. } g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m$$



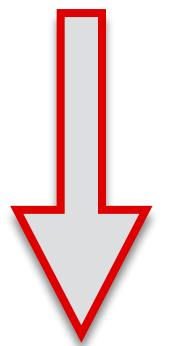
$$\mathcal{L}(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

# Method of Lagrange Multipliers

- Lagrangian for equality-constrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

$$\text{s.t. } g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m$$



$$\mathcal{L}(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

Lagrange multipliers

# Method of Lagrange Multipliers

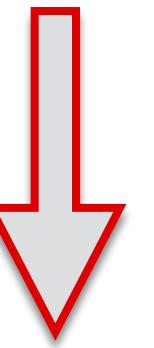
- Solution  $\mathbf{x}^*$  corresponds to stationary point of Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

# Method of Lagrange Multipliers

- Solution  $\mathbf{x}^*$  corresponds to stationary point of Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$



$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = g_i(\mathbf{x}^*) = 0 \quad (\text{feasibility})$$

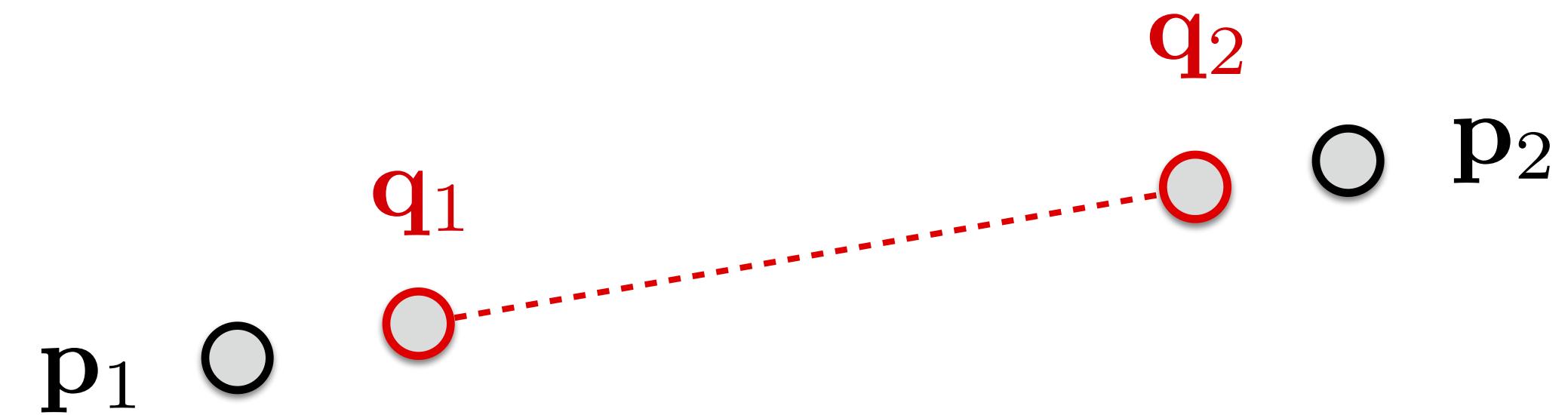
$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \nabla f(\mathbf{x}^*) - \sum_{i=1}^m \lambda_i^* \nabla g_i(\mathbf{x}^*) = \mathbf{0} \quad (\text{linear dependence of gradients})$$

# Method of Lagrange Multipliers

- Useful tool for deriving closed-form solutions
  - Example: given  $\mathbf{p}_1, \dots, \mathbf{p}_m \in \mathbb{R}^3$ , find closest points  $\mathbf{q}_1, \dots, \mathbf{q}_m$  that satisfy certain geometric constraints

# Method of Lagrange Multipliers

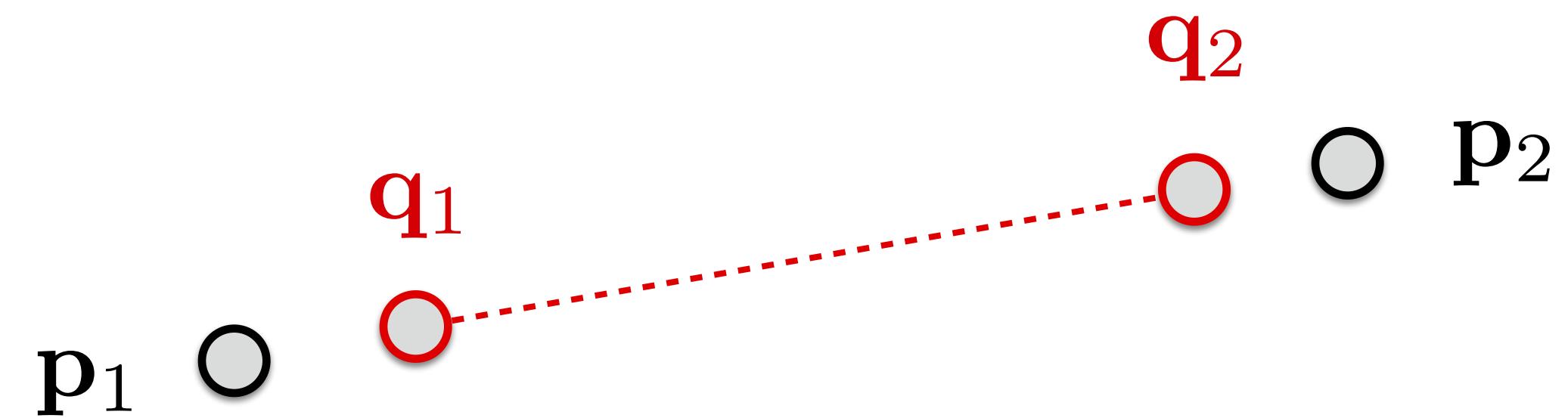
- Given points  $p_1, p_2$ , find closest points  $q_1, q_2$  such that  $\|q_1 - q_2\| = L$



# Method of Lagrange Multipliers

- Optimization problem

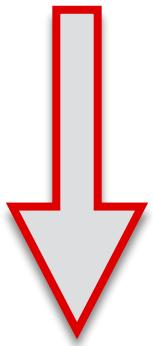
$$\begin{aligned} \min_{\mathbf{q}_1, \mathbf{q}_2} \quad & \|\mathbf{p}_1 - \mathbf{q}_1\|^2 + \|\mathbf{p}_2 - \mathbf{q}_2\|^2 \\ \text{s.t.} \quad & \|\mathbf{q}_1 - \mathbf{q}_2\|^2 = L^2 \end{aligned}$$



# Method of Lagrange Multipliers

- Optimization problem

$$\begin{aligned} \min_{\mathbf{q}_1, \mathbf{q}_2} \quad & \|\mathbf{p}_1 - \mathbf{q}_1\|^2 + \|\mathbf{p}_2 - \mathbf{q}_2\|^2 \\ \text{s.t.} \quad & \|\mathbf{q}_1 - \mathbf{q}_2\|^2 = L^2 \end{aligned}$$



$$\mathcal{L}(\mathbf{q}_1, \mathbf{q}_2, \lambda) = \|\mathbf{p}_1 - \mathbf{q}_1\|^2 + \|\mathbf{p}_2 - \mathbf{q}_2\|^2 - \lambda(\|\mathbf{q}_1 - \mathbf{q}_2\|^2 - L^2)$$

# Method of Lagrange Multipliers

$$\mathcal{L}(\mathbf{q}_1, \mathbf{q}_2, \lambda) = \|\mathbf{p}_1 - \mathbf{q}_1\|^2 + \|\mathbf{p}_2 - \mathbf{q}_2\|^2 - \lambda(\|\mathbf{q}_1 - \mathbf{q}_2\|^2 - L^2)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}_1} = 2[\mathbf{q}_1 - \mathbf{p}_1 - \lambda(\mathbf{q}_1 - \mathbf{q}_2)] = 0$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}_2} = 2[\mathbf{q}_2 - \mathbf{p}_2 - \lambda(\mathbf{q}_2 - \mathbf{q}_1)] = 0$$

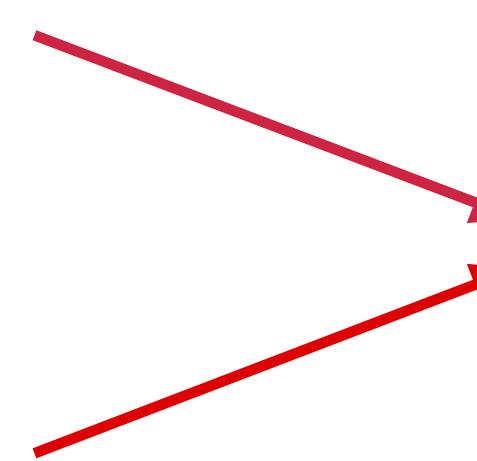
$$\frac{\partial \mathcal{L}}{\partial \lambda} = L^2 - \|\mathbf{q}_1 - \mathbf{q}_2\|^2 = 0$$

# Method of Lagrange Multipliers

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}_1} = 2[\mathbf{q}_1 - \mathbf{p}_1 - \lambda(\mathbf{q}_1 - \mathbf{q}_2)] = 0$$

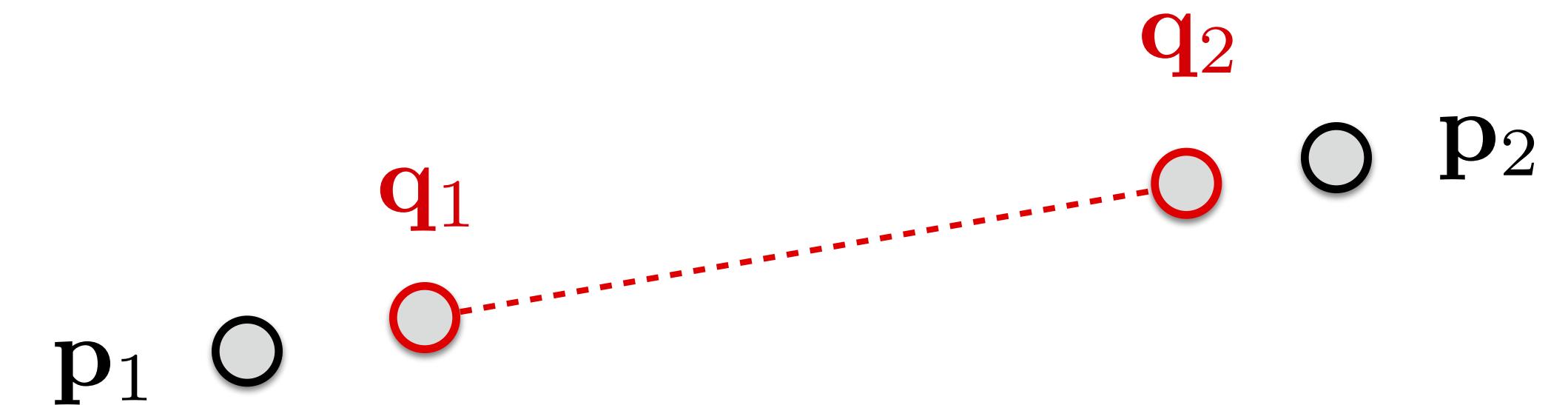
$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}_2} = 2[\mathbf{q}_2 - \mathbf{p}_2 - \lambda(\mathbf{q}_2 - \mathbf{q}_1)] = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = L^2 - \|\mathbf{q}_1 - \mathbf{q}_2\|^2 = 0$$



1.  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{q}_1, \mathbf{q}_2$  colinear

2.  $(\mathbf{p}_1 + \mathbf{p}_2)/2 = (\mathbf{q}_2 + \mathbf{q}_1)/2$



# Numerical Solver

- Many constrained optimization problems have no closed-form solution
  - Method of Lagrange multipliers amounts to solving nonlinear equations: sometimes challenging even for numerical solvers

# Numerical Solver

- Many constrained optimization problems have no closed-form solution
  - Method of Lagrange multipliers amounts to solving nonlinear equations: sometimes challenging even for numerical solvers
  - More sophisticated numerical solvers needed

# Penalty Method

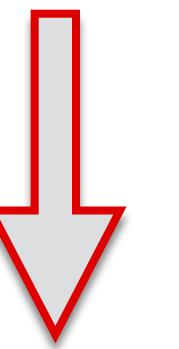
- Convert constrained optimization into unconstrained optimization

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

# Penalty Method

- Convert constrained optimization into unconstrained optimization

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$



$$\min_{\mathbf{x}} f(\mathbf{x}) + \mu \sum_{i=1}^m [g_i(\mathbf{x})]^2$$

# Penalty Method

- Convert constrained optimization into unconstrained optimization

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

$$\min_{\mathbf{x}} f(\mathbf{x}) + \mu \sum_{i=1}^m [g_i(\mathbf{x})]^2$$

Penalizes constraint violation

Penalty parameter

The diagram illustrates the transformation of a constrained optimization problem into an unconstrained one using the penalty method. The original problem is shown at the top, consisting of an objective function  $f(\mathbf{x})$  and equality constraints  $g_i(\mathbf{x}) = 0$  for  $i = 1, \dots, m$ . Below it, the unconstrained problem is shown, where the original objective  $f(\mathbf{x})$  is augmented by a penalty term  $\mu \sum_{i=1}^m [g_i(\mathbf{x})]^2$ . This new term, highlighted with a red dashed box, represents a quadratic measure of constraint violation. A red arrow points from this term to the text "Penalizes constraint violation". Another red arrow points from the scalar  $\mu$  to the text "Penalty parameter", which describes the role of this parameter in controlling the weight of the penalty term.

# Penalty Method

$$\min_{\mathbf{x}} f(\mathbf{x}) + \mu \sum_{i=1}^m [g_i(\mathbf{x})]^2$$

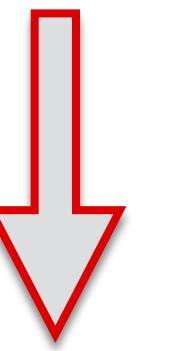
- Can be solved using standard unconstrained solvers
  - unconstrained solution approaches true solution when  $\mu \rightarrow +\infty$
  - ill-conditioned when  $\mu$  becomes too large

# Augmented Lagrangian Method

- Solve another problem with the same solution

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

$$\text{s.t. } g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m$$



$$\min_{\mathbf{x}} f(\mathbf{x}) + \frac{\mu}{2} \sum_{i=1}^m [g_i(\mathbf{x})]^2$$

$$\text{s.t. } g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m$$

# Augmented Lagrangian Method

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) + \frac{\mu}{2} \sum_{i=1}^m [g_i(\mathbf{x})]^2 \\ \text{s.t.} \quad & g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned}$$

- Augmented Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) + \frac{\mu}{2} \sum_{i=1}^m [g_i(\mathbf{x})]^2 - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

# Augmented Lagrangian Method

- Find a saddle point of the augmented Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) + \frac{\mu}{2} \sum_{i=1}^m [g_i(\mathbf{x})]^2 - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

# Augmented Lagrangian Method

- Find a saddle point of the augmented Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) + \frac{\mu}{2} \sum_{i=1}^m [g_i(\mathbf{x})]^2 - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

Iterative solver:

1. Fix  $\{\lambda_i\}$ , minimize  $\mathcal{L}$  over  $\mathbf{x}$ :  $\mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda_1^k, \dots, \lambda_m^k)$
2. Update  $\{\lambda_i\}$ :  $\lambda_i^{k+1} = \lambda_i^k - \mu g_i(\mathbf{x}_i^{k+1})$

# Augmented Lagrangian Method

- Find a saddle point of the augmented Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) + \frac{\mu}{2} \sum_{i=1}^m [g_i(\mathbf{x})]^2 - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

Iterative solver:

1. Fix  $\{\lambda_i\}$ , minimize  $\mathcal{L}$  over  $\mathbf{x}$ :  $\mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda_1^k, \dots, \lambda_m^k)$
2. Update  $\{\lambda_i\}$ :  $\lambda_i^{k+1} = \lambda_i^k - \mu g_i(\mathbf{x}_i^{k+1})$

*Benefit over penalty method:* does not require very large  $\mu$ , avoids ill-conditioning

# General Constrained Optimization

$$\begin{aligned} & \min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \dots, l \end{aligned}$$

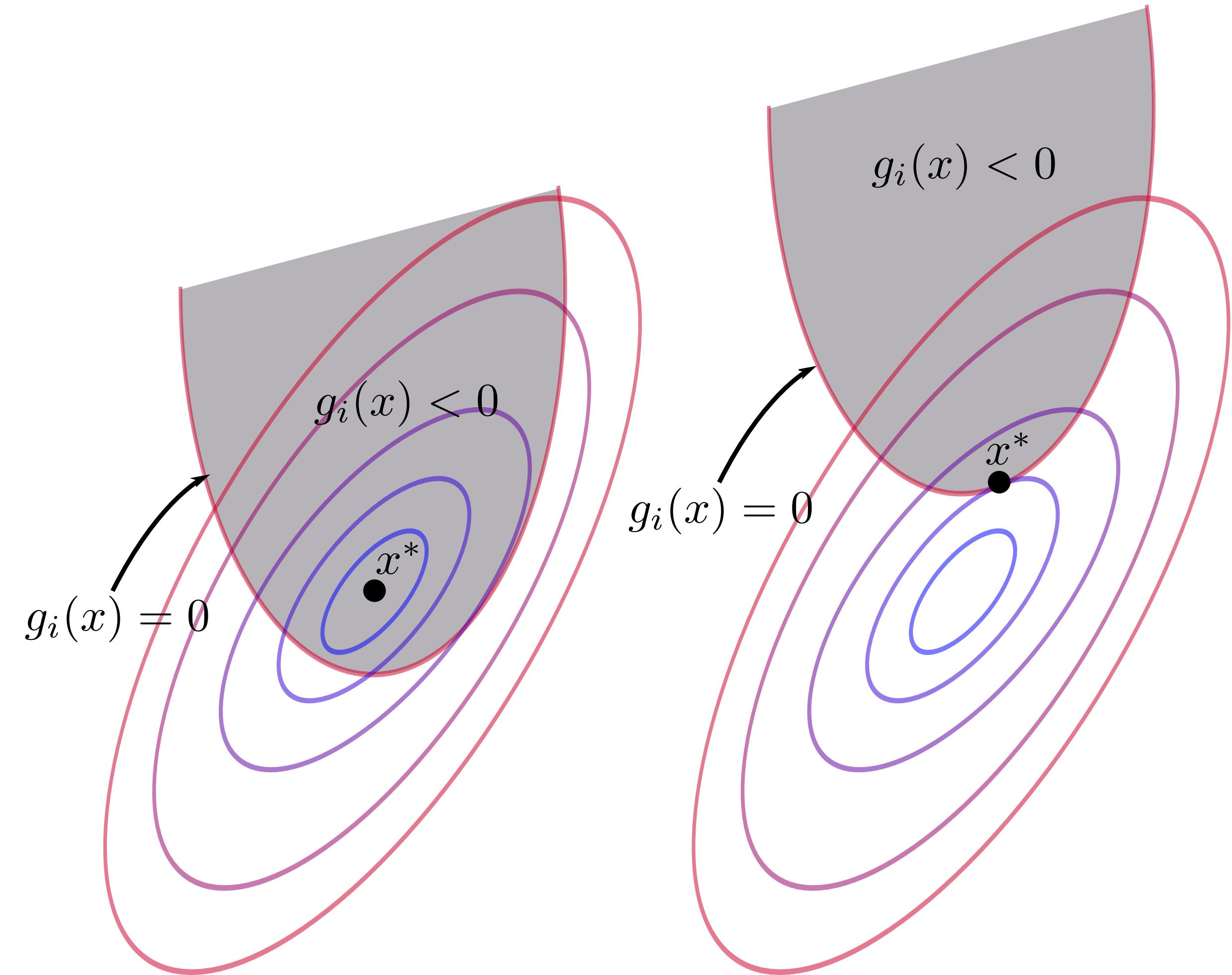
Inequality constraints

Equality constraints

```
graph LR; A[Inequality constraints] --> B["g_i(x) ≤ 0"]; C[Equality constraints] --> D["h_j(x) = 0"]
```

# Inequality Constraints

- Inequality constraints only affect the solution when they becomes equality constraints)



# Karush–Kuhn–Tucker Conditions

## Stationarity

$$-\nabla f(x^*) = \sum_{i=1}^m \mu_i \nabla g_i(x^*) + \sum_{j=1}^{\ell} \lambda_j \nabla h_j(x^*),$$

## Primal feasibility

$$g_i(x^*) \leq 0, \text{ for } i = 1, \dots, m$$

$$h_j(x^*) = 0, \text{ for } j = 1, \dots, \ell$$

## Dual feasibility

$$\mu_i \geq 0, \text{ for } i = 1, \dots, m$$

## Complementary slackness

$$\mu_i g_i(x^*) = 0, \text{ for } i = 1, \dots, m.$$