More on Constrained Optimisation

Yingzhen Li

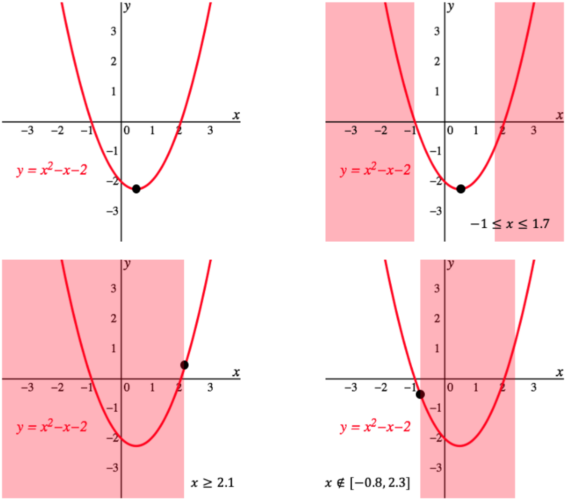
Department of Computing
Imperial College London

@liyzhen2
yingzhen.li@imperial.ac.uk

November 23, 2021

From unconstrained to constrained optimisation

Find minimum for function $f(x) = x^2 - x - 2$:



Constrained optimisation: Set-up

A constrained optimisation problem typically has the following form:

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g_i(\boldsymbol{x}) \leqslant 0, \; i = 1, ..., N \quad \text{(inequality constraints)}$$
$$h_j(\boldsymbol{x}) = 0, \; j = 1, ..., M \quad \text{(equality constraints)}$$

Strategies to solve a constrained problem:

- reparameterise $\boldsymbol{x}$ to directly satisfy the constraints
- **Last lecture**: the Lagrange multiplier method

Constrained optimisation: Set-up

A constrained optimisation problem typically has the following form:

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\begin{aligned} \text{subject to} \quad & g_i(\boldsymbol{x}) \leqslant 0,\ i = 1, ..., N \quad \text{(inequality constraints)} \\ & h_j(\boldsymbol{x}) = 0,\ j = 1, ..., M \quad \text{(equality constraints)} \end{aligned}$$

Strategies to solve a constrained problem:

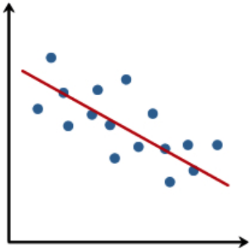
- reparameterise $\boldsymbol{x}$ to directly satisfy the constraints
- **Last lecture**: the Lagrange multiplier method
- **This lecture**: the Karush-Kuhn-Tucker (KKT) conditions

Example: Constrained linear regression

Fitting **constrained** linear regression:

- Dataset: $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$,
  $\mathbf{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_N]^\top \in \mathbb{R}^{N \times D}$,
  $\mathbf{y} = [y_1, ..., y_N]^\top \in \mathbb{R}^{N \times 1}$
- Goal: find $\boldsymbol{\theta} \in \mathbb{R}^{D \times 1}$ such that

$$\mathbf{y} \approx \mathbf{X}\boldsymbol{\theta}$$
$$||\boldsymbol{\theta}||_2^2 \leqslant \delta, \quad \delta > 0$$



Example: Constrained linear regression

A typical **constrained** linear regression model:

- $\boldsymbol{x} \in \mathbb{R}^{D \times 1}$: input features; $y \in \mathbb{R}$: output value
- Model:

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{x}^\top \boldsymbol{\theta}, \quad y = f(\boldsymbol{x}, \boldsymbol{\theta}) + \epsilon, \ \epsilon \sim \mathcal{N}(0, \sigma^2)$$

- The **constrained optimisation** problem:

$$\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \frac{1}{2\sigma^2} ||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2$$

$$\text{subject to} \quad ||\boldsymbol{\theta}||_2^2 - \delta \leqslant 0$$

Example: Constrained linear regression

Solution with Lagrange multiplier method:
Write down an alternative objective called **Lagrangian**:

$$L(\boldsymbol{\theta}; \lambda) = \frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2 + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

- We will find minimum of $L(\boldsymbol{\theta}; \lambda)$ w.r.t. $\boldsymbol{\theta}$
  - as part of the KKT conditions

Example: Constrained linear regression

Solution with Lagrange multiplier method:
Write down an alternative objective called **Lagrangian**:

$$L(\boldsymbol{\theta}; \lambda) = \frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2 + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

- We will find minimum of $L(\boldsymbol{\theta}; \lambda)$ w.r.t. $\boldsymbol{\theta}$
  - as part of the KKT conditions
- We require the Lagrange multiplier $\lambda \geqslant 0$
  - this encourages minimiser of $L(\boldsymbol{\theta}; \lambda)$ to satisfy the constraint
  - as part of the KKT conditions
- We will solve λ using other parts of the KKT conditions

Example: Constrained linear regression

Solution with Lagrange multiplier method:

$$L(\boldsymbol{\theta}; \lambda) = \frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2 + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

Find minimum of $L(\boldsymbol{\theta}; \lambda)$ w.r.t. $\boldsymbol{\theta}$ by setting $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \lambda) = 0$:

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \lambda) = -\frac{1}{\sigma^2}\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + 2\lambda\boldsymbol{\theta} = 0$$

KKT stationarity condition: $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \lambda) = 0$

$$\Rightarrow \boldsymbol{\theta}^*(\lambda) = (\sigma^2\lambda'\mathbf{I} + \mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}, \quad \lambda' = 2\lambda$$

Example: Constrained linear regression

Solution with Lagrange multiplier method:

$$L(\boldsymbol{\theta}; \lambda) = \frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2 + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

Find λ to substitute in $\boldsymbol{\theta}^*(\lambda) = (2\sigma^2\lambda\mathbf{I} + \mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$:
This is done by solving the remaining KKT conditions:

- **primal feasibility:** $||\boldsymbol{\theta}^*(\lambda)||_2^2 - \delta \leqslant 0$

Example: Constrained linear regression

Solution with Lagrange multiplier method:

$$L(\boldsymbol{\theta}; \lambda) = \frac{1}{2\sigma^2} ||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2 + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

Find λ to substitute in $\boldsymbol{\theta}^*(\lambda) = (2\sigma^2 \lambda \mathbf{I} + \mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$:
This is done by solving the remaining KKT conditions:

- **primal feasibility:** $||\boldsymbol{\theta}^*(\lambda)||_2^2 - \delta \leqslant 0$
- **dual feasibility:** $\lambda \geqslant 0$
- **complementary slackness:** $\lambda(||\boldsymbol{\theta}^*(\lambda)||_2^2 - \delta) = 0$

Example: Constrained linear regression

$$L(\boldsymbol{\theta};\lambda) = \underbrace{\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2}_{=L(\boldsymbol{\theta})} + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

dual feasibility: $\lambda \geqslant 0$

- for any $\boldsymbol{\theta}$ s.t. $||\boldsymbol{\theta}||_2^2 \leqslant \delta$

$$\lambda(||\boldsymbol{\theta}||_2^2 - \delta) \leqslant 0 \Rightarrow L(\boldsymbol{\theta}) \geqslant L(\boldsymbol{\theta};\lambda)$$

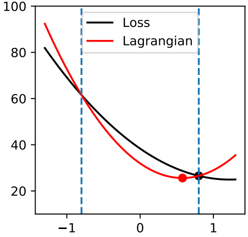
Example: Constrained linear regression

$$L(\boldsymbol{\theta};\lambda) = \underbrace{\frac{1}{2\sigma^2}||\mathbf{y}-\mathbf{X}\boldsymbol{\theta}||_2^2}_{=L(\boldsymbol{\theta})} + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

dual feasibility: $\lambda \geqslant 0$

- for any $\boldsymbol{\theta}$ s.t. $||\boldsymbol{\theta}||_2^2 \leqslant \delta$

$$\lambda(||\boldsymbol{\theta}||_2^2 - \delta) \leqslant 0 \Rightarrow L(\boldsymbol{\theta}) \geqslant L(\boldsymbol{\theta};\lambda)$$



- Assume $\boldsymbol{\theta}_\delta^* = \arg\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta})$,
  then for a fixed $\lambda \geqslant 0$

$$\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}) := L(\boldsymbol{\theta}_\delta^*) \geqslant L(\boldsymbol{\theta}_\delta^*;\lambda) \geqslant \min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta};\lambda)$$

We want a tight lower-bound!

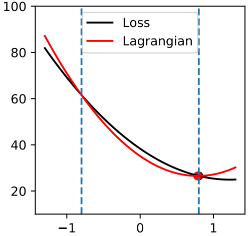
Example: Constrained linear regression

$$L(\boldsymbol{\theta}; \lambda) = \underbrace{\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2}_{=L(\boldsymbol{\theta})} + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

dual feasibility: $\lambda \geqslant 0$

- for any $\boldsymbol{\theta}$ s.t. $||\boldsymbol{\theta}||_2^2 \leqslant \delta$

$$\lambda(||\boldsymbol{\theta}||_2^2 - \delta) \leqslant 0 \Rightarrow L(\boldsymbol{\theta}) \geqslant L(\boldsymbol{\theta}; \lambda)$$



- Assume $\boldsymbol{\theta}_\delta^* = \arg\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta})$, then for a fixed $\lambda \geqslant 0$

$$\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}) := L(\boldsymbol{\theta}_\delta^*) \geqslant L(\boldsymbol{\theta}_\delta^*; \lambda) \geqslant \min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}; \lambda)$$

We want a tight lower-bound!

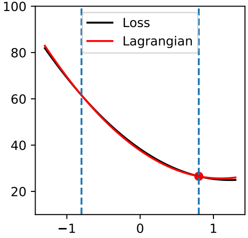
Example: Constrained linear regression

$$L(\boldsymbol{\theta};\lambda) = \underbrace{\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2}_{=L(\boldsymbol{\theta})} + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

dual feasibility: $\lambda \geqslant 0$

- for any $\boldsymbol{\theta}$ s.t. $||\boldsymbol{\theta}||_2^2 \leqslant \delta$

$$\lambda(||\boldsymbol{\theta}||_2^2 - \delta) \leqslant 0 \Rightarrow L(\boldsymbol{\theta}) \geqslant L(\boldsymbol{\theta};\lambda)$$



- Assume $\boldsymbol{\theta}_\delta^* = \arg\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta})$, then for a fixed $\lambda \geqslant 0$

$$\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}) := L(\boldsymbol{\theta}_\delta^*) \geqslant L(\boldsymbol{\theta}_\delta^*;\lambda) \geqslant \min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta};\lambda)$$

We want a tight lower-bound!

Example: Constrained linear regression

$$L(\boldsymbol{\theta}; \lambda) = \underbrace{\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2}_{=L(\boldsymbol{\theta})} + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

complementary slackness: $\lambda(||\boldsymbol{\theta}^*(\lambda)||_2^2 - \delta) = 0$

- For $\lambda^* \geqslant 0$ satisfying complementary slackness, this makes the lower-bound tight:

$$\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}; \lambda^*) = L(\boldsymbol{\theta}^*(\lambda^*); \lambda^*) = L(\boldsymbol{\theta}^*(\lambda^*)) \leqslant \min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta})$$

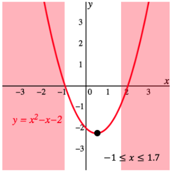
$$\Rightarrow \quad \arg\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}) = \boldsymbol{\theta}^*(\lambda^*)$$

Example: Constrained linear regression

$$L(\boldsymbol{\theta}; \lambda) = \underbrace{\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2}_{=L(\boldsymbol{\theta})} + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

complementary slackness: $\lambda(||\boldsymbol{\theta}^*(\lambda)||_2^2 - \delta) = 0$

- Case 1: $\lambda^* = 0$
  - check **primal feasibility:** $||\boldsymbol{\theta}^*(0)||_2^2 - \delta \leqslant 0$?
  - if so, then $\arg\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}) = \boldsymbol{\theta}^*(0)$
  - in this case we call the constraint **redundant**

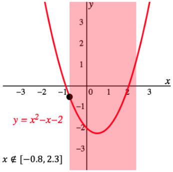


Example: Constrained linear regression

$$L(\boldsymbol{\theta};\lambda) = \underbrace{\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2}_{=L(\boldsymbol{\theta})} + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

complementary slackness: $\lambda(||\boldsymbol{\theta}^*(\lambda)||_2^2 - \delta) = 0$

- Case 2: $\lambda^* > 0$, $||\boldsymbol{\theta}^*(\lambda^*)||_2^2 = \delta$
  - this happens when $||\boldsymbol{\theta}^*(0)||_2^2 - \delta > 0$
  - $\arg\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}) = \boldsymbol{\theta}^*(\lambda^*)$
  - in this case we say the constraint is **binding**

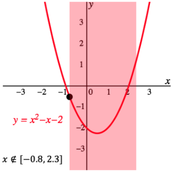


Example: Constrained linear regression

$$L(\boldsymbol{\theta}; \lambda) = \underbrace{\frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2}_{=L(\boldsymbol{\theta})} + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

complementary slackness: $\lambda(||\boldsymbol{\theta}^*(\lambda)||_2^2 - \delta) = 0$

- Case 2: $\lambda^* > 0$, $||\boldsymbol{\theta}^*(\lambda^*)||_2^2 = \delta$
  - this happens when $||\boldsymbol{\theta}^*(0)||_2^2 - \delta > 0$
  - $\arg\min_{||\boldsymbol{\theta}||_2^2 \leqslant \delta} L(\boldsymbol{\theta}) = \boldsymbol{\theta}^*(\lambda^*)$
  - in this case we say the constraint is **binding**



One-to-one relationship between λ^* and δ
by solving $||\boldsymbol{\theta}^*(\lambda)||_2^2 = \delta$ for λ

Connections with ridge regression

Constrained linear regression:

$$\min_{\boldsymbol{\theta}} \frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2$$

$$\text{subject to} \quad ||\boldsymbol{\theta}||_2^2 - \delta \leqslant 0$$

Lagrangian:

$$L(\boldsymbol{\theta}; \lambda) = \frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2 + \lambda(||\boldsymbol{\theta}||_2^2 - \delta)$$

Ridge regression: for a given $\lambda \geqslant 0$

$$\min_{\boldsymbol{\theta}} \frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\boldsymbol{\theta}||_2^2 + \lambda||\boldsymbol{\theta}||_2^2$$

One-to-one relationship: when choosing $\lambda \geqslant 0$ for ridge regression, this is also implicitly choosing δ so that $||\boldsymbol{\theta}^*(\lambda)||_2^2 = \delta$

A summary of the KKT conditions

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$
$$\text{subject to} \quad g_i(\boldsymbol{x}) \leqslant 0,\ i = 1, ..., N$$
$$h_j(\boldsymbol{x}) = 0,\ j = 1, ..., M$$

Lagrangian:

$$L(\boldsymbol{x}; \{\lambda_i\}, \{v_j\}) = L(\boldsymbol{x}) + \sum_{i=1}^{N} \lambda_i g_i(\boldsymbol{x}) + \sum_{j=1}^{M} v_j h_j(\boldsymbol{x})$$

KKT conditions for the solutions of $\boldsymbol{x}, \{\lambda_i\}, \{v_j\}$:

- **stationarity**: $\nabla_x L(\boldsymbol{x}; \{\lambda_i\}, \{v_j\}) = 0$
- **primal feasibility**: $g_i(\boldsymbol{x}) \leqslant 0, h_j(\boldsymbol{x}) = 0$ for all i, j
- **dual feasibility**: $\lambda_i \geqslant 0$ for all i
- **complementary slackness**: $\lambda_i g_i(\boldsymbol{x}) = 0$ for all i

A summary of the KKT conditions

A constrained optimisation problem with equality constraints:

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad h_j(\boldsymbol{x}) = 0, \ j = 1, ..., M \quad \text{(equality constraints)}$$

Lagrangian:

$$L(\boldsymbol{x}; \{v_j\}) = L(\boldsymbol{x}) + \sum_{j=1}^{M} v_j h_j(\boldsymbol{x})$$

KKT conditions for the solutions of $x, \{v_j\}$:

- **stationarity**: $\nabla_x L(\boldsymbol{x}; \{v_j\}) = 0$
- **primal feasibility**: $h_j(\boldsymbol{x}) = 0$ for all j

Dual feasibility & complementary slackness conditions not applicable

(break)

Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

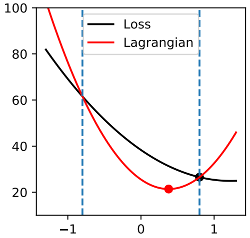
$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

Lagrangian: $L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

dual feasibility: $\lambda \geqslant 0$

- for $\boldsymbol{x}$ satisfying **primal feasibility**:

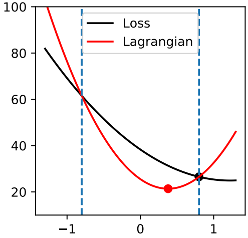
$$\lambda g(\boldsymbol{x}) \leqslant 0, \ \forall \lambda \geqslant 0 \quad \Rightarrow \quad L(\boldsymbol{x}) \geqslant L(\boldsymbol{x}; \lambda)$$



Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$
$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$



Lagrangian: $L(\boldsymbol{x};\lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

dual feasibility: $\lambda \geqslant 0$

- for $\boldsymbol{x}$ satisfying **primal feasibility**:

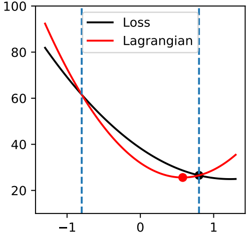
$$\lambda g(\boldsymbol{x}) \leqslant 0,\ \forall \lambda \geqslant 0 \quad \Rightarrow \quad L(\boldsymbol{x}) \geqslant L(\boldsymbol{x};\lambda)$$

$$\min_{g(\boldsymbol{x})\leqslant 0} L(\boldsymbol{x}) := L(\boldsymbol{x}^*) \geqslant L(\boldsymbol{x}^*;\lambda) \geqslant \min_{g(\boldsymbol{x})\leqslant 0} L(\boldsymbol{x};\lambda) \geqslant \underbrace{\min_{\boldsymbol{x}} L(\boldsymbol{x};\lambda)}_{:=L(\boldsymbol{x}^*(\lambda);\lambda)}$$

Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$
$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$



Lagrangian: $L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

dual feasibility: $\lambda \geqslant 0$

- for $\boldsymbol{x}$ satisfying **primal feasibility**:

$$\lambda g(\boldsymbol{x}) \leqslant 0, \ \forall \lambda \geqslant 0 \quad \Rightarrow \quad L(\boldsymbol{x}) \geqslant L(\boldsymbol{x}; \lambda)$$

$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) := L(\boldsymbol{x}^*) \geqslant L(\boldsymbol{x}^*; \lambda) \geqslant \min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}; \lambda) \geqslant \underbrace{\min_{\boldsymbol{x}} L(\boldsymbol{x}; \lambda)}_{:= L(\boldsymbol{x}^*(\lambda); \lambda)}$$

Let's define $D(\lambda) = \min_{\boldsymbol{x}} L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}^*(\lambda); \lambda)$

Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

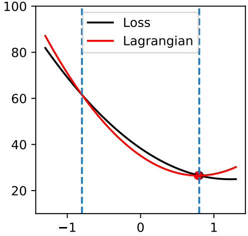
$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

Lagrangian: $L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

dual feasibility: $\lambda \geqslant 0$

- for $\boldsymbol{x}$ satisfying **primal feasibility**:

$$\lambda g(\boldsymbol{x}) \leqslant 0, \ \forall \lambda \geqslant 0 \quad \Rightarrow \quad L(\boldsymbol{x}) \geqslant L(\boldsymbol{x}; \lambda)$$



$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) := L(\boldsymbol{x}^*) \geqslant L(\boldsymbol{x}^*; \lambda) \geqslant D(\lambda) := L(\boldsymbol{x}^*(\lambda); \lambda)$$

The inequalities are true for any $\lambda \geqslant 0$

Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

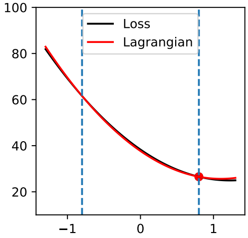
$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

Lagrangian: $L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

dual feasibility: $\lambda \geqslant 0$

- for $\boldsymbol{x}$ satisfying **primal feasibility**:

$$\lambda g(\boldsymbol{x}) \leqslant 0, \ \forall \lambda \geqslant 0 \quad \Rightarrow \quad L(\boldsymbol{x}) \geqslant L(\boldsymbol{x}; \lambda)$$



$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) := L(\boldsymbol{x}^*) \geqslant L(\boldsymbol{x}^*; \lambda^*) \geqslant \underbrace{D(\lambda^*)}_{\lambda^* := \arg\max_{\lambda \geqslant 0} D(\lambda)}$$

Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

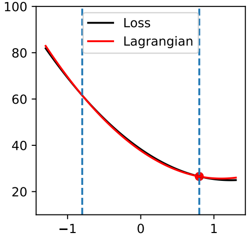
Lagrangian: $L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

dual feasibility: $\lambda \geqslant 0$

- if $\exists\, \lambda^* > 0$ s.t. **complemenary slackness**:

$$\lambda^* g(\boldsymbol{x}^*(\lambda^*)) = 0$$

$$\Rightarrow L(\boldsymbol{x}^*(\lambda^*)) = L(\boldsymbol{x}^*(\lambda^*); \lambda^*)$$



$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) := L(\boldsymbol{x}^*) \geqslant L(\boldsymbol{x}^*; \lambda^*) \geqslant \underbrace{D(\lambda^*)}_{\lambda^* := \arg\max_{\lambda \geqslant 0} D(\lambda)}$$

Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

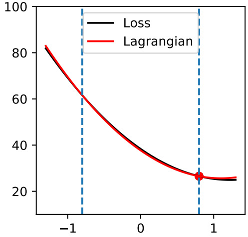
Lagrangian: $L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

dual feasibility: $\lambda \geqslant 0$

- if $\exists\ \lambda^* > 0$ s.t. **complemenary slackness**:

$$\lambda^* g(\boldsymbol{x}^*(\lambda^*)) = 0$$

$$\Rightarrow L(\boldsymbol{x}^*(\lambda^*)) = L(\boldsymbol{x}^*(\lambda^*); \lambda^*)$$



$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) \geqslant \max_{\lambda \geqslant 0} D(\lambda) \geqslant D(\lambda^*) = L(\boldsymbol{x}^*(\lambda^*); \lambda^*) = L(\boldsymbol{x}^*(\lambda^*))$$

Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

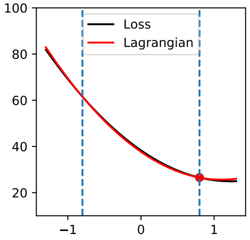
Lagrangian: $L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

dual feasibility: $\lambda \geqslant 0$

- if $\exists\ \lambda^* > 0$ s.t. **complemenary slackness**:

$$\lambda^* g(\boldsymbol{x}^*(\lambda^*)) = 0$$

$$\Rightarrow L(\boldsymbol{x}^*(\lambda^*)) = L(\boldsymbol{x}^*(\lambda^*); \lambda^*)$$



$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) \geqslant \max_{\lambda \geqslant 0} D(\lambda) \geqslant D(\lambda^*) = L(\boldsymbol{x}^*(\lambda^*); \lambda^*) = L(\boldsymbol{x}^*(\lambda^*))$$

Notice $g(\boldsymbol{x}^*(\lambda^*)) \leqslant 0 \quad \Rightarrow \quad \boldsymbol{x}^* = \boldsymbol{x}^*(\lambda^*)$

Dual optimisation

The primal problem

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

Lagrangian: $L(\boldsymbol{x}; \lambda) = L(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

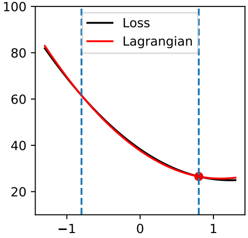
dual feasibility: $\lambda \geqslant 0$

- if $\exists\ \lambda^* > 0$ s.t. **complemenary slackness**:

$$\lambda^* g(\boldsymbol{x}^*(\lambda^*)) = 0$$

$$\Rightarrow L(\boldsymbol{x}^*(\lambda^*)) = L(\boldsymbol{x}^*(\lambda^*); \lambda^*)$$

$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) = \max_{\lambda \geqslant 0} D(\lambda) = D(\lambda^*)$$



Maximise $D(\lambda)$ w.r.t. λ to find such λ^*!

Primal & dual forms

Primal problem:

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

Dual problem:

$$\max_{\lambda} D(\lambda) := \min_{\boldsymbol{x}} (L(\boldsymbol{x}) + \lambda g(\boldsymbol{x}))$$

$$\text{subject to} \quad \lambda \geqslant 0$$

Primal & dual forms

Primal problem:

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

Dual problem:

$$\max_{\lambda} D(\lambda) := \min_{\boldsymbol{x}}(L(\boldsymbol{x}) + \lambda g(\boldsymbol{x}))$$

$$\text{subject to} \quad \lambda \geqslant 0$$

Connected via the KKT conditions:

- $\boldsymbol{x}^*(\lambda)$ and $D(\lambda) = L(\boldsymbol{x}^*(\lambda); \lambda)$ comes from stationarity
- **We assumed primal feasibility** $g(\boldsymbol{x}^*(\lambda^*)) \leqslant 0$
- $\lambda \geqslant 0$ comes from dual feasibility
- **the max. dual problem comes from complementary slackness**

Primal & dual forms

Primal problem:

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

Dual problem:

$$\max_{\lambda} D(\lambda) := \min_{\boldsymbol{x}} (L(\boldsymbol{x}) + \lambda g(\boldsymbol{x}))$$

$$\text{subject to} \quad \lambda \geqslant 0$$

Connected via the KKT conditions:

- $\boldsymbol{x}^*(\lambda)$ and $D(\lambda) = L(\boldsymbol{x}^*(\lambda); \lambda)$ comes from stationarity
- ~~**We assumed primal feasibility** $g(\boldsymbol{x}^*(\lambda^*)) \leqslant 0$~~
- $\lambda \geqslant 0$ comes from dual feasibility
- **the max. dual problem comes from complemetary slackness**

Primal & dual forms

Primal problem:

$$\min_{\boldsymbol{x}} L(\boldsymbol{x})$$

$$\text{subject to} \quad g(\boldsymbol{x}) \leqslant 0$$

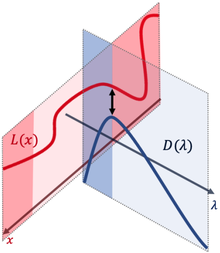
Dual problem:

$$\max_{\lambda} D(\lambda) := \min_{\boldsymbol{x}} (L(\boldsymbol{x}) + \lambda g(\boldsymbol{x}))$$

$$\text{subject to} \quad \lambda \geqslant 0$$

Weak duality (always hold):

$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) \geqslant \max_{\lambda \geqslant 0} D(\lambda)$$



Primal & dual forms

Primal problem:

$$\min L(\boldsymbol{\theta})$$

$$\text{subject to} \quad g(\boldsymbol{\theta}) \leqslant 0$$

Dual problem:

$$\max D(\lambda) := \min_{\boldsymbol{\theta}} (L(\boldsymbol{\theta}) + \lambda g(\boldsymbol{\theta}))$$

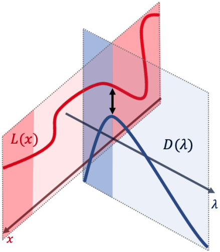
$$\text{subject to} \quad \lambda \geqslant 0$$

Weak duality (always hold):

$$\min_{g(\boldsymbol{x}) \leqslant 0} L(\boldsymbol{x}) \geqslant \max_{\lambda \geqslant 0} D(\lambda)$$

Strong duality

$$\min_{g(\boldsymbol{\theta}) \leqslant 0} L(\boldsymbol{\theta}) = \max_{\lambda \geqslant 0} D(\lambda)$$



Summary

- Constrained optimisation with equality & inequality constraints
  - The Lagrange multiplier method
  - KKT conditions
- Examples
  - PCA (equality constraints)
  - Linear regression with inequality constraints (ℓ_2-norm constraints, connections with ridge regression)
- Primal & dual forms
  - Primal minimisation problem with inequality constraints $\Rightarrow$ dual maximisation problem with inequality constraints
  - Connections shown using KKT conditions
  - Duality gap: Dual problem might only return a lower-bound