Constrained Optimisation

Yingzhen Li

Department of Computing
Imperial College London

@liyzhen2
yingzhen.li@imperial.ac.uk

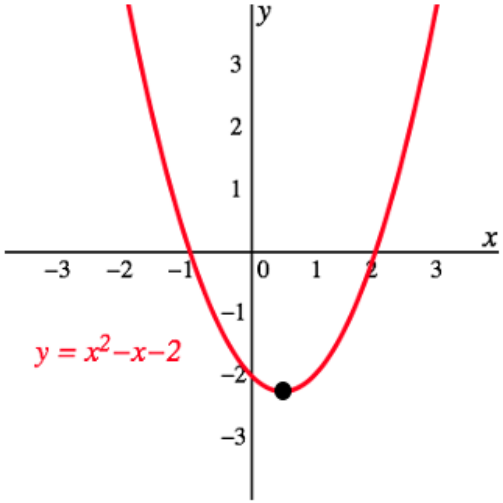
November 19, 2021

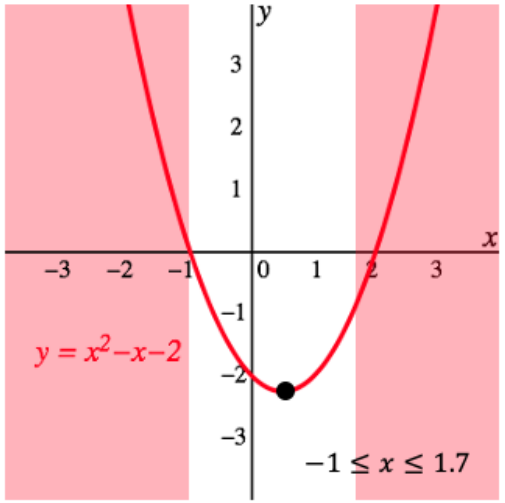
Reading for this week and next week

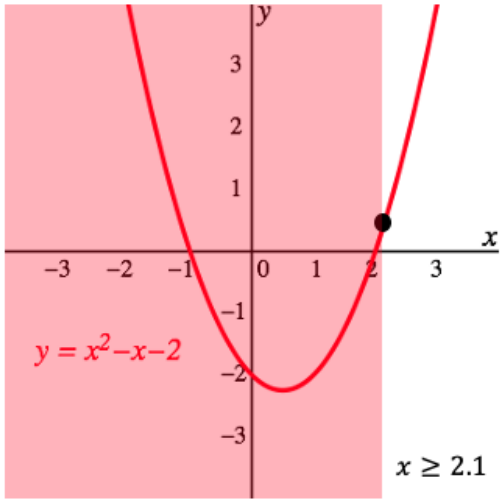
Read MML book: Section 7.2
Extra note will be uploaded to course materials.

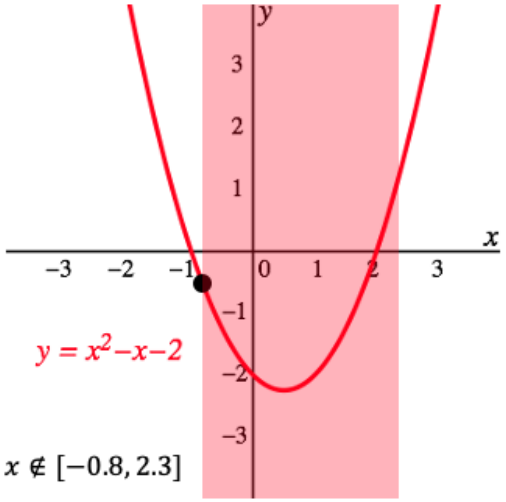
From unconstrained to constrained optimisation

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









Constrained optimisation: Set-up

A constrained optimisation problem typically has the following form:

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

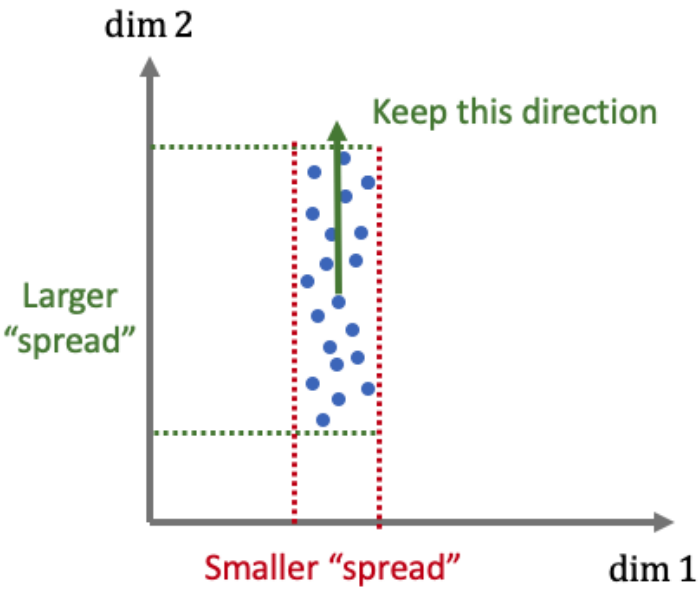
$$\begin{aligned} \text{subject to} \quad & g_i(\boldsymbol{x}) \leqslant 0,\ i = 1, ..., N \quad \text{(inequality constraints)} \\ & h_j(\boldsymbol{x}) = 0,\ j = 1, ..., M \quad \text{(equality constraints)} \end{aligned}$$

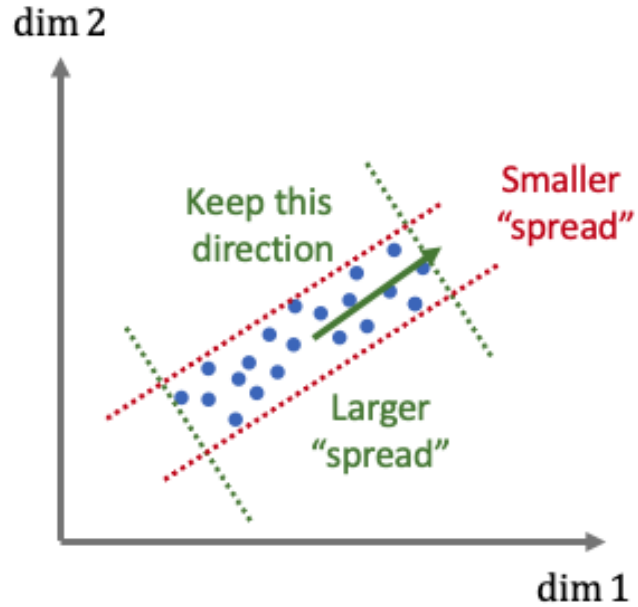
Strategies to solve a constrained problem:

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

PCA: maximum variance perspective

PCA projects data onto directions
where the datapoints "vary the most"





"Spread" is defined as the variance along a given direction

PCA: maximum variance perspective

Recall the problem set-up:

- Data: $\mathcal{D} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^{D\times 1}$ s.t. $\text{mean}(\boldsymbol{x}_n) = \mathbf{0}$
- Find projections in a **lower-dimensional** space:

$$\boldsymbol{z}_n := \mathbf{B}^\top \boldsymbol{x}_n \quad \Leftrightarrow \quad z_{nj} := \mathbf{b}_j^\top \boldsymbol{x}_n$$

using an **orthonormal basis**

$$\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_M], \quad \mathbf{b}_m \in \mathbb{R}^{D\times 1}, \quad M < D$$

PCA: maximum variance perspective

Recall the problem set-up:

- Data: $\mathcal{D} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^{D\times 1}$ s.t. $\text{mean}(\boldsymbol{x}_n) = \mathbf{0}$
- Find projections in a **lower-dimensional** space:

$$\boldsymbol{z}_n := \mathbf{B}^\top \boldsymbol{x}_n \quad \Leftrightarrow \quad z_{nj} := \mathbf{b}_j^\top \boldsymbol{x}_n$$

using an **orthonormal basis**

$$\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_M], \quad \mathbf{b}_m \in \mathbb{R}^{D\times 1}, \quad M < D$$

- Solve for $\mathbf{B}$ such that

$$\sum_{m=1}^{M} \mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n] \quad \text{is maximised}$$

PCA: maximum variance perspective

Solve for the following constrained optimisation problem:

$$\max_{\{\mathbf{b}_1,...,\mathbf{b}_M\}} \sum_{m=1}^{M} \mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$$

$$\text{subject to} \quad ||\mathbf{b}_m||_2^2 = 1,\ m = 1, ..., M$$
$$\mathbf{b}_i \perp \mathbf{b}_j, \quad \forall i \neq j$$

PCA: maximum variance perspective

Solve for the following constrained optimisation problem:

$$\max_{\{\mathbf{b}_1,...,\mathbf{b}_M\}} \sum_{m=1}^{M} \mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n]$$

$$\text{subject to} \quad ||\mathbf{b}_m||_2^2 = 1,\ m = 1, ..., M$$
$$\mathbf{b}_i \perp \mathbf{b}_j, \quad \forall i \neq j$$

- Expanding the objective (see previous lecture on PCA)

$$\mathbb{V}[\mathbf{b}_m^\top \boldsymbol{x}_n] = \mathbf{b}_m^\top \mathbf{S} \mathbf{b}_m, \quad \mathbf{S} := \frac{1}{N} \mathbf{X}\mathbf{X}^\top$$

- Rewrite the maximisation objective with a **minimisation** objective

PCA: maximum variance perspective

Solve for the following constrained optimisation problem:

$$\min_{\{\mathbf{b}_1,...,\mathbf{b}_M\}} -\sum_{m=1}^{M} \mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m$$

$$\text{subject to} \quad ||\mathbf{b}_m||_2^2 = 1,\ m = 1,...,M$$
$$\mathbf{b}_i \perp \mathbf{b}_j, \quad \forall i \neq j$$

PCA: maximum variance perspective

Solve for the following constrained optimisation problem:

$$\min_{\{\mathbf{b}_1,...,\mathbf{b}_M\}} -\sum_{m=1}^{M} \mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m$$

$$\text{subject to} \quad ||\mathbf{b}_m||_2^2 = 1,\ m = 1,...,M$$
$$\mathbf{b}_i \perp \mathbf{b}_j, \quad \forall i \neq j$$

Let's solve it by the Lagrange multiplier method.
Lagrangian with Lagrange multipliers $\{\lambda_m\}$, $\{\gamma_{ij}\}$:

$$L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = -\sum_{m=1}^{M} \mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m + \sum_{m=1}^{M} \lambda_m(||\mathbf{b}_m||_2^2 - 1) + \sum_{i<j} \gamma_{ij}\mathbf{b}_i^\top \mathbf{b}_j$$

PCA: maximum variance perspective

Lagrangian:

$$L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = -\sum_{m=1}^{M} \mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m + \sum_{m=1}^{M} \lambda_m(||\mathbf{b}_m||_2^2 - 1) + \sum_{i<j} \gamma_{ij}\mathbf{b}_i^\top \mathbf{b}_j$$

Solutions of $\{\mathbf{b}_1, ..., \mathbf{b}_M\}$ and the Lagrange multipliers need to satisfy:

- **stationarity**: $\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = \mathbf{0}, m = 1, ..., M$
- **primal feasibility**: $||\mathbf{b}_m||_2^2 = 1, \mathbf{b}_i \perp \mathbf{b}_j$
  (it also means $\nabla_{\lambda_m} L = 0$ and $\nabla_{\gamma_{ij}} L = 0$)

PCA: maximum variance perspective

Lagrangian:

$$L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = -\sum_{m=1}^{M} \mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m + \sum_{m=1}^{M} \lambda_m(||\mathbf{b}_m||_2^2 - 1) + \sum_{i<j} \gamma_{ij}\mathbf{b}_i^\top \mathbf{b}_j$$

Solution of the stationarity condition: make
$\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = \mathbf{0},\ m = 1,...,M$

$$\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = -2\mathbf{S}\mathbf{b}_m + 2\lambda_m\mathbf{b}_m + \sum_{i<m} \gamma_{im}\mathbf{b}_i + \sum_{j>m} \gamma_{mj}\mathbf{b}_j = \mathbf{0}$$

PCA: maximum variance perspective

Lagrangian:

$$L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = -\sum_{m=1}^{M} \mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m + \sum_{m=1}^{M} \lambda_m(||\mathbf{b}_m||_2^2 - 1) + \sum_{i<j} \gamma_{ij}\mathbf{b}_i^\top \mathbf{b}_j$$

Solution of the stationarity condition: make
$\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = \mathbf{0},\ m = 1,...,M$

$$\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\},\{\lambda_m\},\{\gamma_{ij}\}) = -2\mathbf{S}\mathbf{b}_m + 2\lambda_m\mathbf{b}_m + \sum_{i<m}\gamma_{im}\mathbf{b}_i + \sum_{j>m}\gamma_{mj}\mathbf{b}_j = \mathbf{0}$$

$$\Rightarrow \quad \mathbf{b}_m = (-2\mathbf{S} + 2\lambda_m\mathbf{I})^{-1}\left(\sum_{i<m}\gamma_{im}\mathbf{b}_i + \sum_{j>m}\gamma_{mj}\mathbf{b}_j\right)$$

Looks difficult to solve...

PCA: maximum variance perspective

Gradient of the Lagrangian w.r.t. $\mathbf{b}_m$:

$$\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) = -2\mathbf{S}\mathbf{b}_m + 2\lambda_m \mathbf{b}_m + \sum_{i<m} \gamma_{im}\mathbf{b}_i + \sum_{j>m} \gamma_{mj}\mathbf{b}_j = \mathbf{0}$$

Notice the **primal feasibility** condition:

$$\mathbf{b}_m^\top \mathbf{b}_j = 0, \forall j \neq m$$

PCA: maximum variance perspective

Gradient of the Lagrangian w.r.t. $\mathbf{b}_m$:

$$\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) = -2\mathbf{S}\mathbf{b}_m + 2\lambda_m \mathbf{b}_m + \sum_{i<m} \gamma_{im}\mathbf{b}_i + \sum_{j>m} \gamma_{mj}\mathbf{b}_j = \mathbf{0}$$

Notice the **primal feasibility** condition:

$$\mathbf{b}_m^\top \mathbf{b}_j = 0, \forall j \neq m$$

Left-multiply the gradient $\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\})$ with $\mathbf{b}_m^\top$:

$$\mathbf{b}_m^\top \nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) = -2\mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m + 2\lambda_m \mathbf{b}_m^\top \mathbf{b}_m + \underbrace{\sum_{i<m} \gamma_{im}\mathbf{b}_m^\top \mathbf{b}_i + \sum_{j>m} \gamma_{mj}\mathbf{b}_m^\top \mathbf{b}_j}_{=0 \text{ (required by primal feasibility)}}$$

PCA: maximum variance perspective

Gradient of the Lagrangian w.r.t. $\mathbf{b}_m$:

$$\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) = -2\mathbf{S}\mathbf{b}_m + 2\lambda_m \mathbf{b}_m + \sum_{i<m} \gamma_{im}\mathbf{b}_i + \sum_{j>m} \gamma_{mj}\mathbf{b}_j = \mathbf{0}$$

Notice the **primal feasibility** condition:

$$\mathbf{b}_m^\top \mathbf{b}_j = 0, \forall j \neq m$$

Left-multiply the gradient $\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\})$ with $\mathbf{b}_m^\top$:

$$\begin{aligned}\mathbf{b}_m^\top \nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) &= -2\mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m + 2\lambda_m \mathbf{b}_m^\top \mathbf{b}_m \\ &= \underbrace{-2\mathbf{b}_m^\top(\mathbf{S}\mathbf{b}_m - \lambda_m \mathbf{b}_m)}_{=0 \text{ (stationarity condition)}}\end{aligned}$$

PCA: maximum variance perspective

Gradient of the Lagrangian w.r.t. $\mathbf{b}_m$:

$$\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) = -2\mathbf{S}\mathbf{b}_m + 2\lambda_m \mathbf{b}_m + \sum_{i<m} \gamma_{im}\mathbf{b}_i + \sum_{j>m} \gamma_{mj}\mathbf{b}_j = \mathbf{0}$$

Notice the **primal feasibility** condition:

$$\mathbf{b}_m^\top \mathbf{b}_j = 0, \forall j \neq m$$

Left-multiply the gradient $\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\})$ with $\mathbf{b}_m^\top$:

$$\begin{aligned}\mathbf{b}_m^\top \nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) &= -2\mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m + 2\lambda_m \mathbf{b}_m^\top \mathbf{b}_m \\ &= \underbrace{-2\mathbf{b}_m^\top(\mathbf{S}\mathbf{b}_m - \lambda_m \mathbf{b}_m)}_{=0 \text{ (stationarity condition)}}\end{aligned}$$

- $||\mathbf{b}_m||_2^2 = 1$ (primal feasibility) $\Rightarrow$ make $\mathbf{S}\mathbf{b}_m - \lambda_m \mathbf{b}_m = \mathbf{0}$
  $\Rightarrow$ $\mathbf{b}_m$ is an eigenvector of $\mathbf{S}$

PCA: maximum variance perspective

Denote $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$ with $\mathbf{Q} = [\boldsymbol{q}_1, ..., \boldsymbol{q}_D]$.
Notice that for
$\{(\mathbf{b}_1, \lambda_1), ..., (\mathbf{b}_M, \lambda_M)\} \subset \{(\boldsymbol{q}_1, \Lambda_{11}), ..., (\boldsymbol{q}_D, \Lambda_{DD})\}, \gamma_{ij} = 0$
both conditions are satisfied:

- **stationarity**: $\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) = -2(\mathbf{S}\mathbf{b}_m - \lambda_m \mathbf{b}_m) = \mathbf{0}$
- **primal feasibility**: $||\mathbf{b}_m||_2^2 = 1, \mathbf{b}_i \perp \mathbf{b}_j$

PCA: maximum variance perspective

Denote $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$ with $\mathbf{Q} = [\boldsymbol{q}_1, ..., \boldsymbol{q}_D]$.
Notice that for
$\{(\mathbf{b}_1, \lambda_1), ..., (\mathbf{b}_M, \lambda_M)\} \subset \{(\boldsymbol{q}_1, \Lambda_{11}), ..., (\boldsymbol{q}_D, \Lambda_{DD})\}, \gamma_{ij} = 0$
both conditions are satisfied:

- **stationarity**: $\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) = -2(\mathbf{S}\mathbf{b}_m - \lambda_m \mathbf{b}_m) = \mathbf{0}$
- **primal feasibility**: $||\mathbf{b}_m||_2^2 = 1, \mathbf{b}_i \perp \mathbf{b}_j$

$\Rightarrow$ Solutions from the Lagrangian provide **local minima** only.

PCA: maximum variance perspective

Denote $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^\top$ with $\mathbf{Q} = [\boldsymbol{q}_1, ..., \boldsymbol{q}_D]$.
Notice that for
$\{(\mathbf{b}_1, \lambda_1), ..., (\mathbf{b}_M, \lambda_M)\} \subset \{(\boldsymbol{q}_1, \Lambda_{11}), ..., (\boldsymbol{q}_D, \Lambda_{DD})\}, \gamma_{ij} = 0$
both conditions are satisfied:

- **stationarity**: $\nabla_{\mathbf{b}_m} L(\{\mathbf{b}_m\}, \{\lambda_m\}, \{\gamma_{ij}\}) = -2(\mathbf{S}\mathbf{b}_m - \lambda_m \mathbf{b}_m) = \mathbf{0}$
- **primal feasibility**: $||\mathbf{b}_m||_2^2 = 1, \mathbf{b}_i \perp \mathbf{b}_j$

$\Rightarrow$ Solutions from the Lagrangian provide **local minima** only.

Get the **global minimum** by choosing the best from all local optima:

$$min \sum_{m=1}^{M} -\mathbf{b}_m^\top \mathbf{S}\mathbf{b}_m \quad \text{for } \{\mathbf{b}_1, ..., \mathbf{b}_M\} \subset \{\boldsymbol{q}_1, ..., \boldsymbol{q}_D\}$$

$$\Rightarrow \quad \{\mathbf{b}_1, ..., \mathbf{b}_M\} \text{ is the leading } M \text{ eigenvectors of } \mathbf{S}$$

Constrained optimisation: equality constraints

A constrained optimisation problem with equality constraints:

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

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

The Lagrange multiplier method:

Constrained optimisation: equality constraints

A constrained optimisation problem with equality constraints:

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

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

The Lagrange multiplier method:

- Construct the Lagrangian function

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

Constrained optimisation: equality constraints

A constrained optimisation problem with equality constraints:

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

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

The Lagrange multiplier method:

- Construct the Lagrangian function

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

- Find stationary points for the Lagrangian: solve for $\underbrace{\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \{\lambda_j\}) = \mathbf{0}}_{\text{stationarity}}$ and $\underbrace{\nabla_{\lambda_j} L(\boldsymbol{x}, \{\lambda_j\}) = 0, j = 1, ..., M}_{\text{primal feasibility}}$

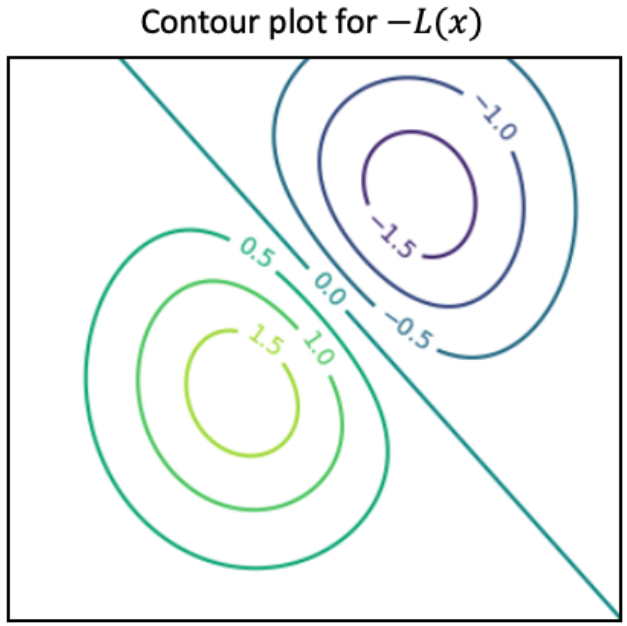
Intuition for the Lagrange multiplier method

A simplified set-up: $M = 1, \lambda := \lambda_1, h(\boldsymbol{x}) := h_1(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^D$

$$L(x, \lambda) = L(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

Intuition for solving $\nabla_{x,\lambda} L = \mathbf{0}$:

- A "surface" (i.e. low-dim manifold) in $\mathbb{R}^D$ can be represented as $f(\boldsymbol{x}) = c$
  - **Contour plot:** plotting different manifolds $f(\boldsymbol{x}) = c_i$ with different c_i
- ...so we can plot the contour of $-L(\boldsymbol{x})$



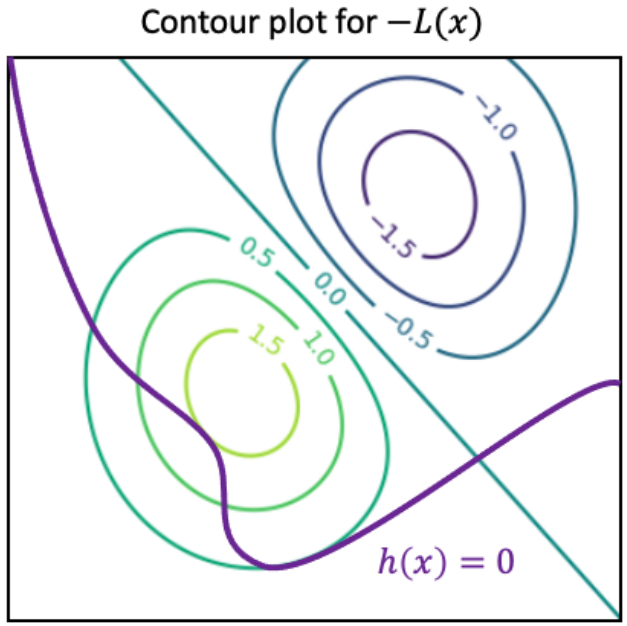
Intuition for the Lagrange multiplier method

A simplified set-up: $M = 1, \lambda := \lambda_1, h(\boldsymbol{x}) := h_1(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^D$

$$L(x, \lambda) = L(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

Intuition for solving $\nabla_{x,\lambda} L = \mathbf{0}$:

- A "surface" (i.e. low-dim manifold) in $\mathbb{R}^D$ can be represented as $f(\boldsymbol{x}) = c$
  - **Constraint:** $h(\boldsymbol{x}) = 0$ represents a low-dim manifold in $\mathbb{R}^D$
- ...so we can plot the "surface" of $h(\boldsymbol{x}) = 0$



Intuition for the Lagrange multiplier method

A simplified set-up: $M = 1, \lambda := \lambda_1, h(\boldsymbol{x}) := h_1(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^D$

$$L(x, \lambda) = L(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

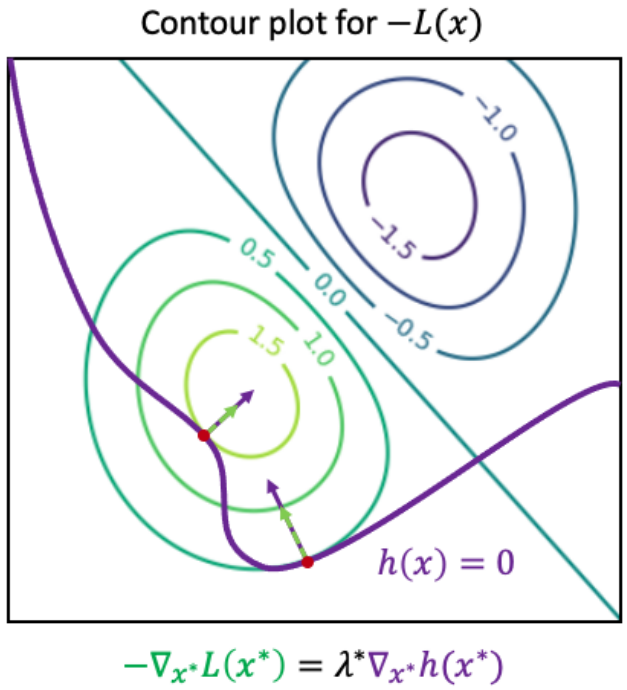
Intuition for solving $\nabla_{x,\lambda} L = \mathbf{0}$:

- A "surface" (i.e. low-dim manifold) in $\mathbb{R}^D$ can be represented as $f(\boldsymbol{x}) = c$
- Solving $\nabla_{x,\lambda} L = \mathbf{0}$:

$$\Rightarrow \quad -\nabla_{x^*} L(\boldsymbol{x}^*) = \lambda^* \nabla_{x^*} h(\boldsymbol{x}^*)$$

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

$\Rightarrow$ For some c^*, $-L(\boldsymbol{x}) = c^*$ and $h(\boldsymbol{x}) = 0$ are **tangent** at point $\boldsymbol{x}^*$



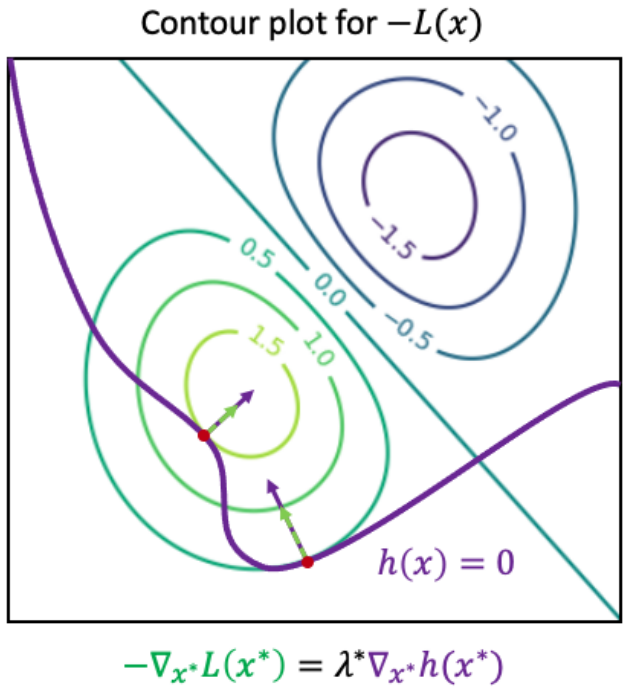
Intuition for the Lagrange multiplier method

A simplified set-up: $M = 1, \lambda := \lambda_1, h(\boldsymbol{x}) := h_1(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^D$

$$L(x, \lambda) = L(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

Intuition for solving $\nabla_{x,\lambda} L = \mathbf{0}$:

- Why finding **"points of contact"**:
  - If can find $\delta \boldsymbol{x}$ s.t. $h(\boldsymbol{x} + \delta \boldsymbol{x}) = 0$,
    $\langle -\nabla_{\boldsymbol{x}} L(\boldsymbol{x}), \delta \boldsymbol{x} \rangle > 0$
    $\Rightarrow L(\boldsymbol{x}) > L(\boldsymbol{x} + \delta \boldsymbol{x})$



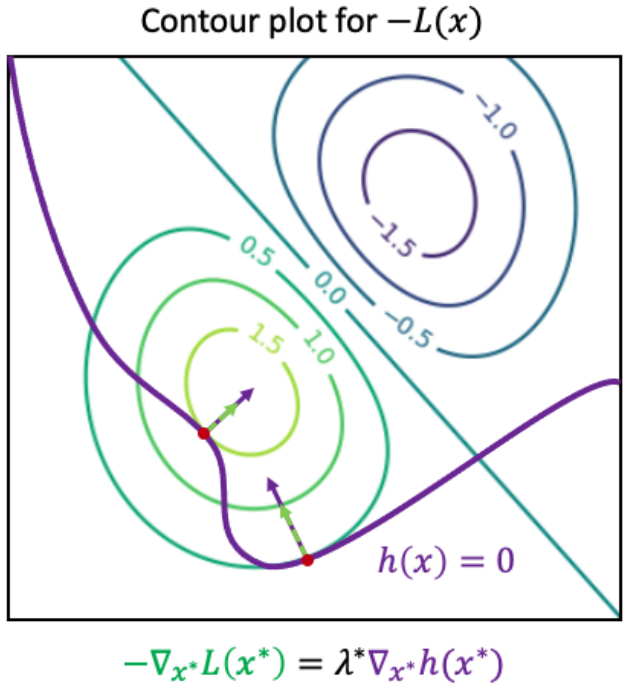
Intuition for the Lagrange multiplier method

A simplified set-up: $M = 1, \lambda := \lambda_1, h(\boldsymbol{x}) := h_1(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^D$

$$L(x, \lambda) = L(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

Intuition for solving $\nabla_{x,\lambda} L = \mathbf{0}$:

- Why finding **"points of contact"**:
  - If can find $\delta \boldsymbol{x}$ s.t. $h(\boldsymbol{x} + \delta \boldsymbol{x}) = 0$,
    $\langle -\nabla_{\boldsymbol{x}} L(\boldsymbol{x}), \delta \boldsymbol{x} \rangle > 0$
    $\Rightarrow L(\boldsymbol{x}) > L(\boldsymbol{x} + \delta \boldsymbol{x})$
  - To make $\boldsymbol{x}^*$ a local minimum of $L(\boldsymbol{x})$ on surface $h(\boldsymbol{x}) = 0$:
    make sure $\forall \delta \boldsymbol{x}$ s.t. $h(\boldsymbol{x} + \delta \boldsymbol{x}) = 0$,
    $\langle -\nabla_{\boldsymbol{x}} L(\boldsymbol{x}), \delta \boldsymbol{x} \rangle \leqslant 0$
    $\Rightarrow$ $\boldsymbol{x}^*$ is a "point of contact"



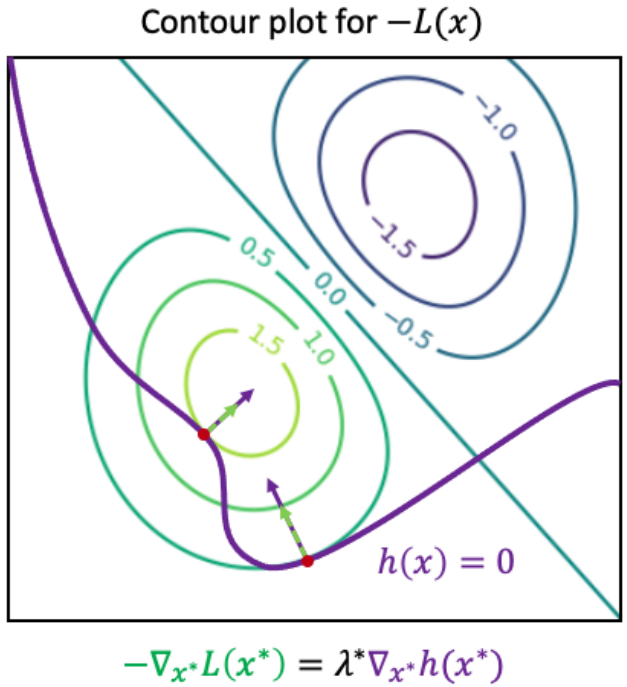
Intuition for the Lagrange multiplier method

A simplified set-up: $M = 1, \lambda := \lambda_1, h(\boldsymbol{x}) := h_1(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^D$

$$L(x, \lambda) = L(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

Intuition for solving $\nabla_{x,\lambda} L = \mathbf{0}$:

- Why finding **"points of contact"**:
  - $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$ is a normal vector of surface $f(\boldsymbol{x}) = c$ at $\boldsymbol{x}$
  - Find "points of contact" between surfaces $-L(\boldsymbol{x}) = c^*$ and $h(\boldsymbol{x}) = 0$: $\Leftrightarrow$ find $\boldsymbol{x}^*$ s.t. $-\nabla_{\boldsymbol{x}^*} L(\boldsymbol{x}^*)$ and $\nabla_{\boldsymbol{x}^*} h(\boldsymbol{x}^*)$ align on the same line



Summary

Summary for today's lecture:

- Constrained optimisation with **equality** constraints
  - The Lagrange multiplier method
  - PCA (maximum variance perspective)
- Exercise for today:
  - formulate PCA (minimum reconstruction error perspective) as constrained optimisation
  - solve it with the Lagrange multiplier method

Summary

Summary for today's lecture:

- Constrained optimisation with **equality** constraints
  - The Lagrange multiplier method
  - PCA (maximum variance perspective)
- Exercise for today:
  - formulate PCA (minimum reconstruction error perspective) as constrained optimisation
  - solve it with the Lagrange multiplier method

Next lecture:

- Constrained optimisation with **inequality** constraints
  - The KKT conditions
  - Ridge regression: constrained optimisation perspective