Data processing and networks optimization

Part I: Introduction

Pierre Borgnat[1], **Jean-Christophe Pesquet**[2], **Nelly Pustelnik**[1]

[1] ENS Lyon – Laboratoire de Physique – CNRS UMR 5672
pierre.borgnat@ens-lyon.fr, nelly.pustelnik@ens-lyon.fr

[2] LIGM – Univ. Paris-Est – CNRS UMR 8049
jean-christophe.pesquet@univ-paris-est.fr

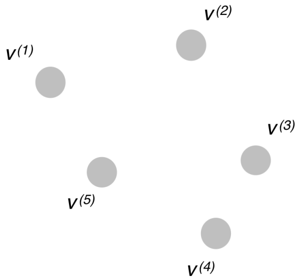
Optimization ?

Whatever people do, at some point they get a craving to organize things in a best possible way. This intention, converted in a mathematical form, turns out to be an optimization problem of certain type.

(Yurii Nesterov)

Valued graphs

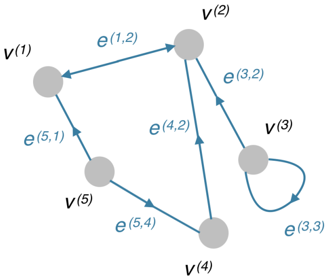
- $\mathcal{V} = \{v^{(i)} \mid i \in \{1, \ldots, M\}\}$
  $\rightsquigarrow$ set of vertices $=$ objects
  $v^{(i)} \in \mathcal{V} \leftrightarrow i \in \{1, \ldots, M\}$



Valued graphs

- $\mathcal{V} = \{v^{(i)} \mid i \in \{1, \ldots, M\}\}$
  $\rightsquigarrow$ set of vertices $=$ objects
  $v^{(i)} \in \mathcal{V} \leftrightarrow i \in \{1, \ldots, M\}$

- $\mathcal{E} = \{e^{(i,j)} \mid (i,j) \in \mathbb{E}\}$
  $\rightsquigarrow$ set of edges $=$ object relationships
  $e^{(i,j)} \in \mathcal{E} \leftrightarrow (i,j) \in \mathbb{E}$

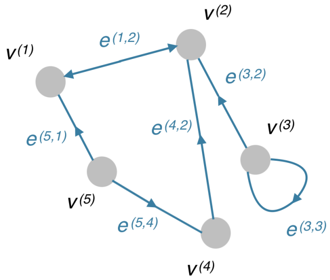


Valued graphs

- $\mathcal{V} = \{v^{(i)} \mid i \in \{1, \ldots, M\}\}$
  $\rightsquigarrow$ set of vertices $=$ objects
  $v^{(i)} \in \mathcal{V} \leftrightarrow i \in \{1, \ldots, M\}$

- $\mathcal{E} = \{e^{(i,j)} \mid (i,j) \in \mathbb{E}\}$
  $\rightsquigarrow$ set of edges $=$ object relationships
  $e^{(i,j)} \in \mathcal{E} \leftrightarrow (i,j) \in \mathbb{E}$

- directed reflexive graph: $|\mathbb{E}| \leq M^2$

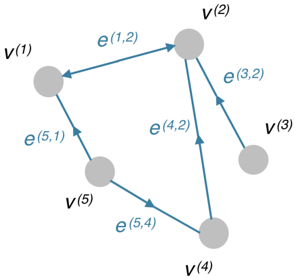


Valued graphs

- $\mathcal{V} = \left\{ v^{(i)} \mid i \in \{1, \ldots, M\} \right\}$
  $\rightsquigarrow$ set of vertices = objects
  $v^{(i)} \in \mathcal{V} \leftrightarrow i \in \{1, \ldots, M\}$

- $\mathcal{E} = \left\{ e^{(i,j)} \mid (i,j) \in \mathbb{E} \right\}$
  $\rightsquigarrow$ set of edges = object relationships
  $e^{(i,j)} \in \mathcal{E} \leftrightarrow (i,j) \in \mathbb{E}$

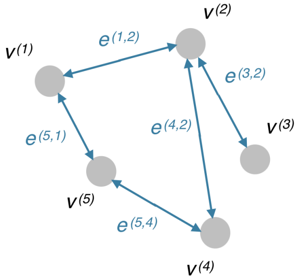
- directed reflexive graph: $|\mathbb{E}| \leq M^2$
  directed nonreflexive graph: $|\mathbb{E}| \leq M(M-1)$



Valued graphs

- $\mathcal{V} = \{v^{(i)} \mid i \in \{1, \ldots, M\}\}$
  $\rightsquigarrow$ set of vertices $=$ objects
  $v^{(i)} \in \mathcal{V} \leftrightarrow i \in \{1, \ldots, M\}$

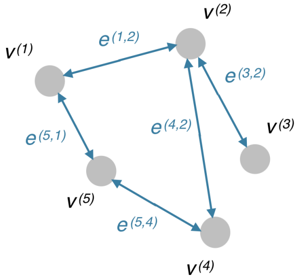
- $\mathcal{E} = \{e^{(i,j)} \mid (i,j) \in \mathbb{E}\}$
  $\rightsquigarrow$ set of edges $=$ object relationships
  $e^{(i,j)} \in \mathcal{E} \leftrightarrow (i,j) \in \mathbb{E}$



- directed reflexive graph: $|\mathbb{E}| \leq M^2$
  directed nonreflexive graph: $|\mathbb{E}| \leq M(M-1)$
  undirected nonreflexive graph: $|\mathbb{E}| \leq M(M-1)/2$

Valued graphs

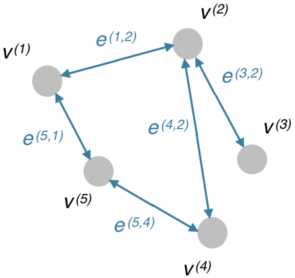
- $\mathcal{V} = \{v^{(i)} \mid i \in \{1, \ldots, M\}\}$
  $\rightsquigarrow$ set of vertices = objects
  $v^{(i)} \in \mathcal{V} \leftrightarrow i \in \{1, \ldots, M\}$



- $\mathcal{E} = \{e^{(i,j)} \mid (i,j) \in \mathbb{E}\}$
  $\rightsquigarrow$ set of edges = object relationships
  $e^{(i,j)} \in \mathcal{E} \leftrightarrow (i,j) \in \mathbb{E}$
- directed reflexive graph: $|\mathbb{E}| \leq M^2$
  directed nonreflexive graph: $|\mathbb{E}| \leq M(M-1)$
  undirected nonreflexive graph: $|\mathbb{E}| \leq M(M-1)/2$
- $(x^{(i)})_{1 \leq i \leq M}$: weights on vertices (scalars or vectors)

Valued graphs

- $\mathcal{V} = \left\{ v^{(i)} \mid i \in \{1, \ldots, M\} \right\}$
  $\rightsquigarrow$ set of vertices = objects
  $v^{(i)} \in \mathcal{V} \leftrightarrow i \in \{1, \ldots, M\}$



- $\mathcal{E} = \left\{ e^{(i,j)} \mid (i,j) \in \mathbb{E} \right\}$
  $\rightsquigarrow$ set of edges = object relationships
  $e^{(i,j)} \in \mathcal{E} \leftrightarrow (i,j) \in \mathbb{E}$
- directed reflexive graph: $|\mathbb{E}| \leq M^2$
  directed nonreflexive graph: $|\mathbb{E}| \leq M(M-1)$
  undirected nonreflexive graph: $|\mathbb{E}| \leq M(M-1)/2$
- $(x^{(i)})_{1 \leq i \leq M}$: weights on vertices (scalars or vectors)
- $(x^{(i,j)})_{(i,j) \in \mathbb{E}}$: weights on edges (scalars or vectors)

Quantitative analysis

- Objective function
  The merits of a given choice of the weights is evaluated by

$$f\left((x^{(i)})_{1\le i\le M}, (\mathrm{x}^{(i,j)})_{(i,j)\in\mathbb{E}}\right)$$

Quantitative analysis

- **Objective function**

  The merits of a given choice of the weights is evaluated by

  $$f\Big(\underbrace{(x^{(i)})_{1\le i\le M}, (\mathrm{x}^{(i,j)})_{(i,j)\in\mathbb{E}}}_{x}\Big)$$

  where

  $$x = \begin{bmatrix} (x^{(i)})_{1\le i\le M} \\ (\mathrm{x}^{(i,j)})_{(i,j)\in\mathbb{E}} \end{bmatrix} \in D \subset \mathbb{R}^N$$

  and $f\colon D \mapsto [-\infty, +\infty]$.

  Example: scalar weights $\Rightarrow N = M + |\mathbb{E}|$

  The number of variables can often be reduced.

Optimization over graphs

- Minimization problems

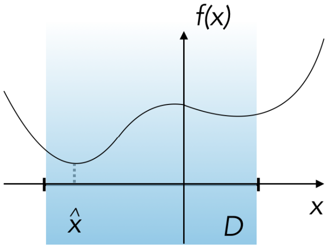
  f: cost function

  We want to

$$\begin{aligned} & \text{Find } \widehat{x} \in D \quad \text{such that} \quad (\forall x \in D)\ f(\widehat{x}) \leq f(x) \\ \Leftrightarrow \quad & \text{Find } \widehat{x} \in D \quad \text{such that} \quad f(\widehat{x}) = \inf_{x \in D} f(x) \end{aligned}$$

  that is

$$\text{Find } \widehat{x} \in \operatorname*{Argmin}_{x \in D} \; f(x).$$

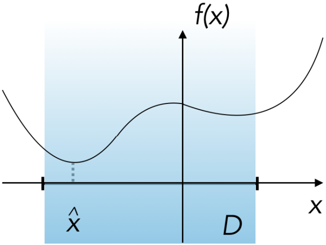


Optimization over graphs

- Maximization problems
  f: reward function
  We want to

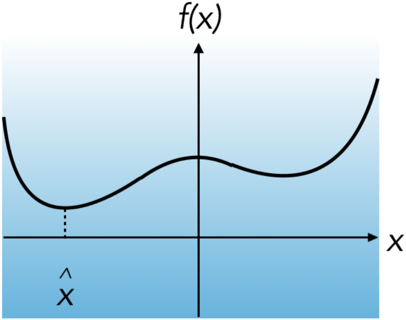
$$
\begin{aligned}
& \text{Find } \widehat{x} \in D \quad \text{such that} \quad (\forall x \in D)\ f(\widehat{x}) \geq f(x) \\
\Leftrightarrow \quad & \text{Find } \widehat{x} \in D \quad \text{such that} \quad (\forall x \in D)\ -f(\widehat{x}) \leq -f(x) \\
\Leftrightarrow \quad & \text{Find } \widehat{x} \in \underset{x \in D}{\mathrm{Argmin}} \ \big(-f(x)\big).
\end{aligned}
$$

Without loss of generality, we can focus on minimization problems with $f\colon D \to \,]-\infty, +\infty]$.



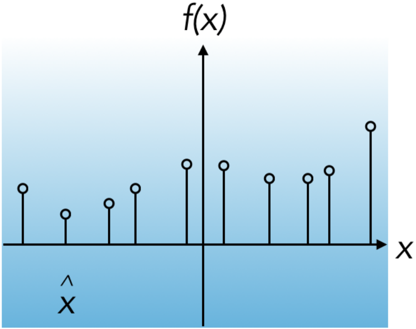
Various types of minimization problems

- $D = \mathbb{R}^N$: unconstrained problem



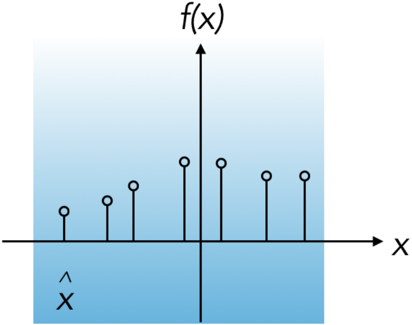
Various types of minimization problems

- $D = \mathbb{R}^N$: unconstrained problem
- D countable: discrete optimization problem



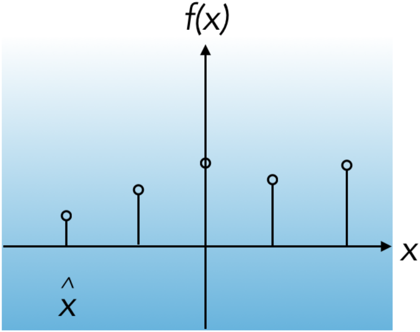
Various types of minimization problems

- $D = \mathbb{R}^N$: unconstrained problem
- D countable: discrete optimization problem
  - D finite: combinatorial optimization problem



Various types of minimization problems

- $D = \mathbb{R}^N$: unconstrained problem
- D countable: discrete optimization problem
  - D finite: combinatorial optimization problem
  - $D \subset \mathbb{Z}^N$: integer optimization problem



Various types of minimization problems

- D uncountable: continuous optimization problem
- Optimization problem with P equality constraints and Q inequality constraints:

$$\begin{aligned} D = \{x \in \mathbb{R}^N \mid (\forall i \in \{1,\ldots,P\})\ &\varphi_i(x) = \delta_i \\ \text{and } (\forall j \in \{1,\ldots,Q\})\ &\psi_j(x) \le \eta_j\} \end{aligned}$$

where $(\forall i \in \{1,\ldots,P\})\ \delta_i \in \mathbb{R}$ and $\varphi_i \colon \mathbb{R}^N \to \,]-\infty,+\infty]$,
$(\forall j \in \{1,\ldots,Q\})\ \eta_j \in \mathbb{R}$ and $\psi_j \colon \mathbb{R}^N \to \,]-\infty,+\infty]$.
If $\varphi_i \colon x \mapsto \langle x \mid u_i \rangle$ with $i \in \{1,\ldots,P\}$ and $u_i \in \mathbb{R}^N$, then *linear* (or affine) equality constraint.
If $\psi_j \colon x \mapsto \langle x \mid u_j \rangle$ with $j \in \{1,\ldots,Q\}$ and $u_j \in \mathbb{R}^N$, then *linear* (or affine) inequality constraint.

Various types of minimization problems

Remark:

$$
\begin{aligned}
& \text{Find } \widehat{x} \in \underset{x \in D}{\operatorname{Argmin}} \; f(x) \\
\Leftrightarrow \quad & \text{Find } \widehat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \; \widetilde{f}(x)
\end{aligned}
$$

where

$$
(\forall x \in \mathbb{R}^N) \quad \widetilde{f}(x) = \begin{cases} f(x) & \text{if } x \in D \\ +\infty & \text{otherwise.} \end{cases}
$$

Allowing non finite valued functions leads to a unifying view of constrained and unconstrained minimization problems.

Main questions to be addressed

1. Existence/uniqueness of a solution $\widehat{x}$?

Main questions to be addressed

1. Existence/uniqueness of a solution $\widehat{x}$?
2. Characterization of solutions: necessary/sufficient conditions for $\widehat{x}$ to be a solution.

Main questions to be addressed

1. Existence/uniqueness of a solution $\widehat{x}$?
2. Characterization of solutions: necessary/sufficient conditions for $\widehat{x}$ to be a solution.
3. Designing an algorithm to approximate a solution in the frequent case when no closed form solution is available, i.e. building a sequence $(x_n)_{n\in\mathbb{N}}$ of $\mathbb{R}^N$ such that

$$\lim_{n\to+\infty} x_n = \widehat{x}.$$

Main questions to be addressed

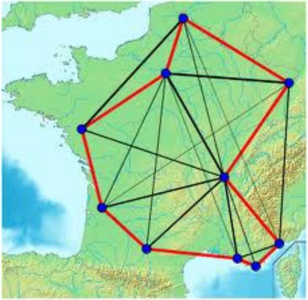
1. Existence/uniqueness of a solution $\widehat{x}$?
2. Characterization of solutions: necessary/sufficient conditions for $\widehat{x}$ to be a solution.
3. Designing an algorithm to approximate a solution in the frequent case when no closed form solution is available, i.e. building a sequence $(x_n)_{n\in\mathbb{N}}$ of $\mathbb{R}^N$ such that

$$\lim_{n\to+\infty} x_n = \widehat{x}.$$

4. Evaluation of the performance of the optimization algorithm:
   - Convergence speed
     Example: If there exists $\rho \in]0,1[$ and $n^* \in \mathbb{N}$ such that $(\forall n \geq n^*)$ $\|x_{n+1} - \widehat{x}\| \leq \rho\|x_n - \widehat{x}\|$, then *Q-linear* convergence rate.
     If $\lim_{n\to+\infty} \frac{\|x_{n+1}-\widehat{x}\|}{\|x_n-\widehat{x}\|} = 0$, then *Q-superlinear* convergence rate.
   - Robustness to numerical errors
   - Amenability to parallel/distributed implementations.

Example: Traveling salesman problem

- Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city?



Example: Traveling salesman problem

- Directed nonreflexive graph

  $\mathcal{V}$: set of the cities

  $\mathcal{E}$: set of roads

  $x = (\mathrm{x}^{(i,j)})_{(i,j)\in\mathbb{E}}$ where, for every $(i,j) \in \mathbb{E}$, $\mathrm{x}^{(i,j)} = 1$ if the path goes from the city i to the city j and 0 otherwise,

  $N = |\mathbb{E}|$.

- Cost function

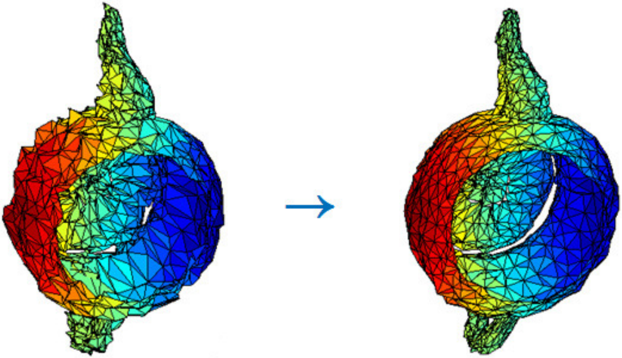
$$f(x) = \sum_{(i,j)\in\mathbb{E}} c_{i,j}\mathrm{x}^{(i,j)} \quad \text{to be minimized subject to}$$

$$\begin{cases} (\forall (i,j) \in \mathbb{E}) \ \ \mathrm{x}^{(i,j)} \in \{0,1\}, & c_{i,j}\text{: distance between } i \text{ and } j \\ (\forall j \in \{1,\ldots,M\}) \ \ \sum_{i=1,(i,j)\in\mathbb{E}}^{M} \mathrm{x}^{(i,j)} = 1 & \leftarrow \ \text{There is 1 arrival to each city.} \\ (\forall i \in \{1,\ldots,M\}) \ \ \sum_{j=1,(i,j)\in\mathbb{E}}^{M} \mathrm{x}^{(i,j)} = 1 & \leftarrow \ \text{From each city, there is 1 departure.} \end{cases}$$

$\rightsquigarrow D \subset \{0,1\}^N$

Example: Mesh denoising problem

- Remove uncertainties in mesh measurements



Example: Mesh denoising problem

- Undirected nonreflexive graph
  $\mathcal{V}$: set of vertices of the mesh
  $\mathcal{E}$: set of edges of the mesh
  $x = (x^{(i)})_{1\leq i\leq M}$ where, for every $i \in \{1,\ldots,M\}$,
  $x^{(i)} \in \mathbb{R}^3$: 3D coordinates of the i-th vertex of the true object
  $N = 3M$ and $D = \mathbb{R}^N$.

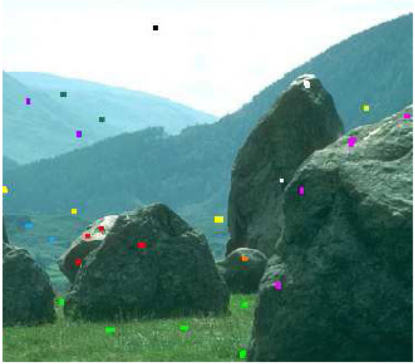
- Cost function

$$f(x) = \sum_{i=1}^{M} \Big(\|x^{(i)} - y^{(i)}\|^2 + \big(\sum_{j\in\mathcal{N}_i} \omega_{i,j} \|x^{(j)} - x^{(i)}\|^2 \big)^{1/2} \Big)$$

where, for every $i \in \{1,\ldots,M\}$,
  - $y^{(i)}$: 3D measured coordinates of the i-th vertex
  - $\mathcal{N}_i$: neighborhood of i-th vertex
  - $(\omega_{i,j})_{j\in\mathcal{N}_i}$: nonnegative regularization constants.

Example: Segmentation problem

- Assign a label to every pixel in an image so that pixels with the same label share certain characteristics.



Example: Segmentation problem

- Undirected nonreflexive regular graph
  $\mathcal{V}$: Set of pixels
  $\mathcal{E}$: Set of neighborhood relationships between pixels
  $x = (x^{(i)})_{1 \le i \le M}$ where, for every $i \in \{1, \ldots, M\}$, $x^{(i)} \in \mathcal{L}$: label of pixel i
  $\mathcal{L} \subset \mathbb{R}$: finite set of labels
  $N = M$ and $D = \mathcal{L}^M$.

- Cost function

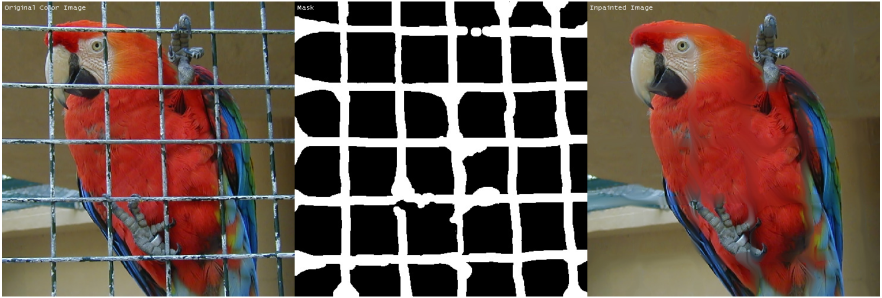
$$f(x) = \sum_{i=1}^{M} \alpha_i^q |x^{(i)} - y^{(i)}|^q + \sum_{(i,j) \in \mathbb{E}} \omega_{i,j}^q |x^{(j)} - x^{(i)}|^q$$

where, for every $i \in \{1, \ldots, M\}$,

- $y^{(i)}$: intensity value of pixel i
- $(\alpha_i)_{1 \le i \le M}$ and $(\omega_{i,j})_{(i,j) \in \mathbb{E}}$: nonnegative constants
- $(p, q) \in [1, +\infty[^2$.

Example: Inpainting

- Given an incomplete image, fill its missing part in a visually plausible way



Example: Inpainting

- Undirected nonreflexive graph

  $\mathcal{V}$: Set of patches intersecting the target region (black area)

  $\mathcal{E}$: Set of neighborhood relationships between these patches

  $x = (x^{(i)})_{1 \leq i \leq M}$ where, for every $i \in \{1, \dots, M\}$, $x^{(i)} \in \mathcal{L}$: patch i

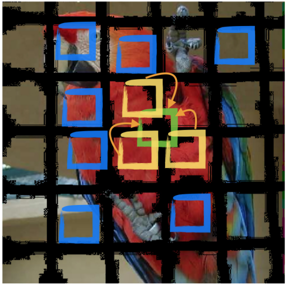
  $\mathcal{L}$: Set of complete patches from the source image (blue boxes)

  $N = M$ and $D = \mathcal{L}^N$.

- Cost function

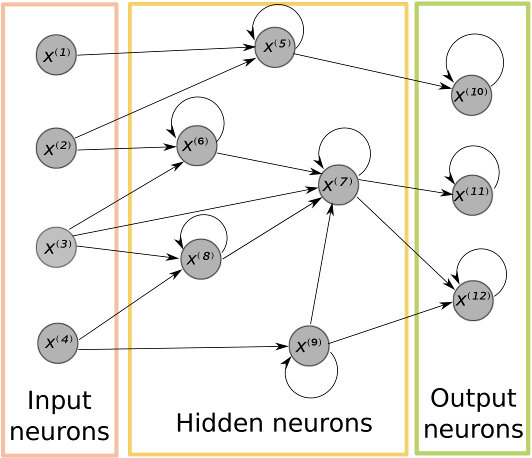
$$f(x) = \sum_{i=1}^{M} U_i(x^{(i)}) + \sum_{(i,j) \in \mathbb{E}} V_{i,j}(x^{(i)}, x^{(j)})$$

  - $(U_i)_{1 \leq i \leq M}$: measure the similarity between the target patch and the source patches
  - $(V_{i,j})_{1 \leq i,j \leq N}$: measure the similarity on overlapping regions.



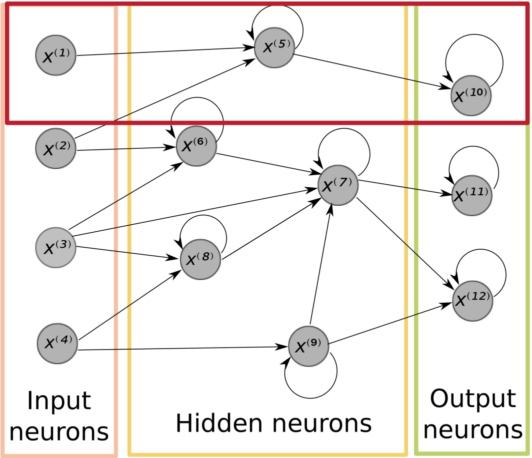
Example: Neural networks

- Mimic biological neural networks



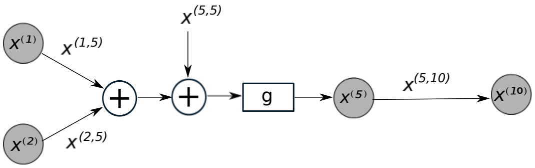
Example: Neural networks

- Mimic biological neural networks



Example: Neural networks

- Mimic biological neural networks



Example: Neural networks

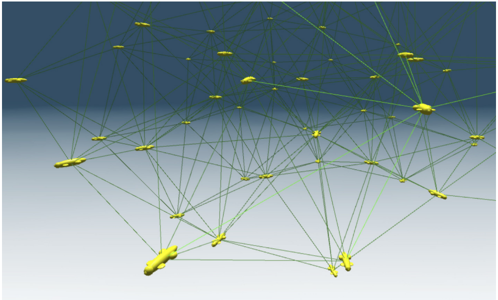
- Directed reflexive graph
  $\mathcal{V}$: Set of neurons in the network
  $\mathcal{E}$: Set of synaptic connections
  $x = (\mathrm{x}^{(i,j)})_{(i,j)\in\mathbb{E}}$ where, for every $(i,j) \in \mathbb{E}$, if $i \neq j$, $\mathrm{x}^{(i,j)}$ is a weighting input factor, and if $i = j$, $\mathrm{x}^{(i,i)}$ is a bias parameter.
  $N = |\mathbb{E}|$ and $D = \mathbb{R}^N$.

- Cost function
$$f(x) = \sum_{i\in\mathbb{V}_O} |x^{(i)} - s^{(i)}|^p, \quad p \in [1, +\infty]$$
  where $(\forall i \in \{1, \ldots, M\})\ x^{(i)} = g(\sum_{j\in\mathcal{N}_i} \mathrm{x}^{(j,i)} x^{(j)} + \mathrm{x}^{(i,i)})$ with $g \colon \mathbb{R} \to \mathbb{R}$ activation function (ex: sigmoid $g = \tanh$), $\mathcal{N}_i$ set of neurons exciting neuron i, $s^{(i)}$ desired response when the neuron is trained with an element $(x^{(j)})_{j\in\mathbb{V}_I}$ of an input training sequence, $\{1, \ldots, M\} = \mathbb{V}_I \cup \mathbb{V}_T \cup \mathbb{V}_O$ with $\mathbb{V}_I$ input layer, $\mathbb{V}_T$ intermediate layers, and $\mathbb{V}_O$ output layer.
  $\rightsquigarrow$ deep learning requiring a large training sequence for the parameters to be estimated in a reliable manner.

Example: Multiagent consensus problem

- Each agent process local data and cooperate with its neighbors to estimate global parameters.

  Examples: Trajectory planning, distributed clustering, Internet traffic anomalies, ...



Example: Multiagent consensus problem

- Directed nonreflexive graph
  $\mathcal{V}$: Set of agents
  $\mathcal{E}$: Set of relationships between agents
  $x = (x^{(i)})_{1 \leq i \leq M}$ vector of estimates in $\mathbb{R}^P$ at each node
  $N = MP$ and

$$D = \left\{(x^{(i)})_{1 \leq i \leq M} \in (\mathbb{R}^P)^M \;\middle|\; x^{(1)} = \ldots = x^{(M)}\right\}.$$

- Cost function

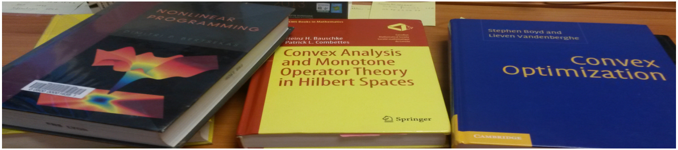
$$f(x) = \sum_{i=1}^{M} g_i(x^{(i)}, y^{(i)})$$

where $(\forall i \in \{1, \ldots, M\})\ -g_i\colon \mathbb{R}^P \times \mathbb{R}^Q \to]-\infty, +\infty]$ utility function at node i and $y^{(i)} \in \mathbb{R}^Q$: observations available at node i.

Optimization ?

- Convex optimization ($\sim$ 8 hours)
  - Fundamentals
  - Dual formulation
  - Non-smooth optimization
  - Algorithms
- Discrete optimization ($\sim$ 3 hours)
  - Graph-cut problem
  - Alpha expansion
- Distributed algorithms ($\sim$ 3 hours)

Reference books



- **D. Bertsekas**, Nonlinear programming, Athena Scientic, Belmont, Massachussets, 1995.
- **Y. Nesterov**, Introductory Lectures on Convex Optimization: A Basic Course, Springer, 2004.
- **S. Boyd and L. Vandenberghe**, Convex optimization, Cambridge University Press, 2004.
- **H. H. Bauschke and P. L. Combettes**, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer, New York, 2011.