Optimization methods
Lecture 10: Intro to combinatorial optimization problems and its convex relaxations

Alexandr Katrutsa

Modern State of Artificial Intelligence Masters Program
Moscow Institute of Physics and Technology

Brief reminder of the previous lecture

- Newton method

Brief reminder of the previous lecture

- Newton method
- Quasi-Newton methods

Brief reminder of the previous lecture

- Newton method
- Quasi-Newton methods
- Limited memory quasi-Newton methods

Gradient-free methods

Why do they need?

- Target vector is discrete
- Gradient is difficult to compute

Gradient-free methods

Why do they need?

- Target vector is discrete
- Gradient is difficult to compute

Examples

- Problems from the decision making and selection of element from the finite set
- Hyper-parameter selection in the machine learning models

Simulated annealing

- Main steps
  - Initialization of the initial vector and parameters
  - Every iteration follows the rule

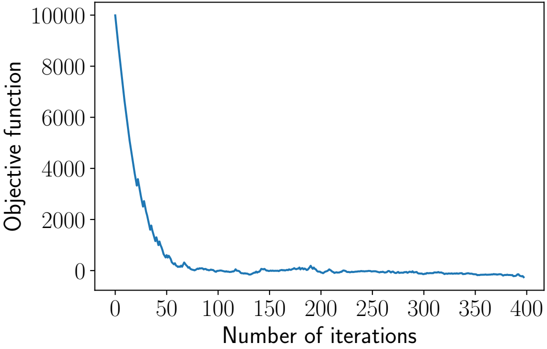
$$\mathbb{P}(\mathbf{x}_k \to \mathbf{x}^*) = \begin{cases} 1 & f(\mathbf{x}^*) < f(\mathbf{x}_k) \\ \exp\left(-\frac{f(\mathbf{x}^*) - f(\mathbf{x}_k)}{T/k}\right) & f(\mathbf{x}^*) > f(\mathbf{x}_k) \end{cases}$$

  - Denominator tuning is heuristic

Example

Partition problem with adjacency matrix $\mathbf{W}$

$$\min \mathbf{x}^\top \mathbf{W} \mathbf{x}$$
$$\text{s.t. } x_i \in \{-1, 1\}$$



$$\alpha_k = 1/k$$

Other gradient-free methods

- Genetic algorithms
- Particle swarm optimization approach
- Many others, more details in the webinar

From LP to SDP

LP in standard form

$$\min_{\mathbf{x}} \langle \mathbf{c}, \mathbf{x} \rangle$$
$$\text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\mathbf{x} \geq 0$$

- Replace vectors by matrices
- $\mathbb{R}^n_+ \to \mathbf{S}^n_+$
- Inner product between vectors $\to$ inner product between matrices

SDP

- Standard form

$$
\begin{aligned}
&\min_{\mathbf{X}} \operatorname{trace}(\mathbf{C}\mathbf{X}) \\
\text{s.t. } &\operatorname{trace}(\mathbf{A}_i\mathbf{X}) = b_i \\
&\mathbf{X} \succeq 0,
\end{aligned}
$$

where $\mathbf{C} \in \mathbf{S}^n$ and $\mathbf{A}_i \in \mathbf{S}^n$

SDP

- Standard form

$$\begin{aligned} \min_{\mathbf{X}} \ & \mathrm{trace}(\mathbf{C}\mathbf{X}) \\ \text{s.t. } & \mathrm{trace}(\mathbf{A}_i\mathbf{X}) = b_i \\ & \mathbf{X} \succeq 0, \end{aligned}$$

  where $\mathbf{C} \in \mathbf{S}^n$ and $\mathbf{A}_i \in \mathbf{S}^n$

- Dual form

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^\top \mathbf{x} \\ \text{s.t. } & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{G} + \sum_{i=1}^{n} x_i \mathbf{F}_i \preceq 0, \end{aligned}$$

  where $\mathbf{G} \in \mathbf{S}^n$ and $\mathbf{F}_i \in \mathbf{S}^n$.

Convex relaxations of the non-convex QP problems

- Initial non-convex QP problem

$$\begin{aligned} &\min \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } &\mathbf{x}^\top \mathbf{A}_i \mathbf{x} + \mathbf{b}_i^\top \mathbf{x} \leq 0 \end{aligned}$$

Convex relaxations of the non-convex QP problems

- Initial non-convex QP problem

$$\begin{aligned} \min \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } \mathbf{x}^\top \mathbf{A}_i \mathbf{x} + \mathbf{b}_i^\top \mathbf{x} \le 0 \end{aligned}$$

- Equivalent transformation

$$\begin{aligned} \min \operatorname{trace}(\mathbf{A}\mathbf{X}) + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } \operatorname{trace}(\mathbf{A}_i \mathbf{X}) + \mathbf{b}_i^\top \mathbf{x} \le 0 \\ \mathbf{X} = \mathbf{x}\mathbf{x}^\top \end{aligned}$$

Convex relaxations of the non-convex QP problems

- Initial non-convex QP problem

$$\begin{aligned} \min \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } \mathbf{x}^\top \mathbf{A}_i \mathbf{x} + \mathbf{b}_i^\top \mathbf{x} \le 0 \end{aligned}$$

- Equivalent transformation

$$\begin{aligned} \min \operatorname{trace}(\mathbf{A}\mathbf{X}) + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } \operatorname{trace}(\mathbf{A}_i \mathbf{X}) + \mathbf{b}_i^\top \mathbf{x} \le 0 \\ \mathbf{X} = \mathbf{x}\mathbf{x}^\top \end{aligned}$$

- Relax non-convex rank constraint

$$\begin{aligned} \min_{\mathbf{X},\mathbf{x}} \operatorname{trace}(\mathbf{A}\mathbf{X}) + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } \operatorname{trace}(\mathbf{A}_i \mathbf{X}) + \mathbf{b}_i^\top \mathbf{x} \le 0 \\ \mathbf{X} - \mathbf{x}\mathbf{x}^\top \succeq 0 \end{aligned}$$

Convex relaxations of the non-convex QP problems

- Initial non-convex QP problem

$$\begin{aligned} \min \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } \mathbf{x}^\top \mathbf{A}_i \mathbf{x} + \mathbf{b}_i^\top \mathbf{x} \le 0 \end{aligned}$$

- Equivalent transformation

$$\begin{aligned} \min \operatorname{trace}(\mathbf{A}\mathbf{X}) + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } \operatorname{trace}(\mathbf{A}_i\mathbf{X}) + \mathbf{b}_i^\top \mathbf{x} \le 0 \\ \mathbf{X} = \mathbf{x}\mathbf{x}^\top \end{aligned}$$

- Relax non-convex rank constraint

$$\begin{aligned} \min_{\mathbf{X},\mathbf{x}} \operatorname{trace}(\mathbf{A}\mathbf{X}) + \mathbf{b}^\top \mathbf{x} \\ \text{s.t. } \operatorname{trace}(\mathbf{A}_i\mathbf{X}) + \mathbf{b}_i^\top \mathbf{x} \le 0 \\ \mathbf{X} - \mathbf{x}\mathbf{x}^\top \succeq 0 \end{aligned}$$

- The last constraint is equivalent to $\begin{bmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^\top & 1 \end{bmatrix} \succeq 0$

MAXCUT

- Given unoriented graph $G = (V, E)$ with the weight matrix $\mathbf{W} \in \mathbf{S}^n$ and $\mathbf{W} \geq 0$

MAXCUT

- Given unoriented graph $G = (V, E)$ with the weight matrix $\mathbf{W} \in \mathbf{S}^n$ and $\mathbf{W} \geq 0$
- We need to find such subset of vertices $S \subset V$, that the sum of vertices weights that connect vertices from S with vertices from $V \setminus S$ **is maximal**

MAXCUT

- Given unoriented graph $G = (V, E)$ with the weight matrix $\mathbf{W} \in \mathbf{S}^n$ and $\mathbf{W} \geq 0$
- We need to find such subset of vertices $S \subset V$, that the sum of vertices weights that connect vertices from S with vertices from $V \setminus S$ **is maximal**

Formal problem statement

$$\max_{\mathbf{x}} \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - x_i x_j)$$

$$\text{s.t. } x_i \in \{+1, \ -1\}$$

Denote by c^* the optimal value of the objective function

SDP relaxation

- Introduce matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$

SDP relaxation

- Introduce matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$
- Then

$$
\begin{aligned}
&\max_{\mathbf{x}} \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - x_{ij}) \\
\text{s.t. } & x_i \in \{+1,\ -1\} \\
& \mathbf{X} = \mathbf{x}\mathbf{x}^\top
\end{aligned}
$$

SDP relaxation

- Introduce matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$
- Then

$$
\begin{aligned}
&\max_{\mathbf{x}} \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - x_{ij}) \\
\text{s.t. } & x_i \in \{+1,\ -1\} \\
& \mathbf{X} = \mathbf{x}\mathbf{x}^\top
\end{aligned}
$$

- Re-write the first constraint in the form $\mathrm{diag}(\mathbf{X}) = \mathbf{1}$

SDP relaxation

- Introduce matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$
- Then

$$
\begin{aligned}
&\max_{\mathbf{x}} \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - x_{ij}) \\
\text{s.t. } & x_i \in \{+1,\ -1\} \\
& \mathbf{X} = \mathbf{x}\mathbf{x}^\top
\end{aligned}
$$

- Re-write the first constraint in the form $\mathrm{diag}(\mathbf{X}) = \mathbf{1}$
- Rank constraint is replaced with $\mathbf{X} \in \mathbf{S}_+^n$

SDP relaxation

- Introduce matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$
- Then

$$\begin{aligned} \max_{\mathbf{x}} \; & \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - x_{ij}) \\ \text{s.t. } & x_i \in \{+1, \; -1\} \\ & \mathbf{X} = \mathbf{x}\mathbf{x}^\top \end{aligned}$$

- Re-write the first constraint in the form $\operatorname{diag}(\mathbf{X}) = \mathbf{1}$
- Rank constraint is replaced with $\mathbf{X} \in \mathbf{S}_+^n$
- Optimal value of the objective function is denoted by p^*

SDP relaxation

- Introduce matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$
- Then

$$\begin{aligned} &\max_{\mathbf{x}} \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - x_{ij}) \\ \text{s.t. } & x_i \in \{+1,\ -1\} \\ & \mathbf{X} = \mathbf{x}\mathbf{x}^\top \end{aligned}$$

- Re-write the first constraint in the form $\mathrm{diag}(\mathbf{X}) = \mathbf{1}$
- Rank constraint is replaced with $\mathbf{X} \in \mathbf{S}_+^n$
- Optimal value of the objective function is denoted by p^*
- Since the feasible set is expanded, then $p^* \geq c^*$

How to reconstruct the solution?

- After solving the relaxed problem we obtain a matrix $\mathbf{X}^*$
  Q: how to recover the cut?

How to reconstruct the solution?

- After solving the relaxed problem we obtain a matrix $\mathbf{X}^*$
  Q: how to recover the cut?

Goemans-Williamson's algorithm

1. Generate random vector $\mathbf{v}$ in the unit sphere
2. $S = \{i \mid \langle \mathbf{v}, \mathbf{u}_i \rangle \geq 0\}$, where $\mathbf{u}_i$ form matrix $\mathbf{U} : \mathbf{X} = \mathbf{U}^\top \mathbf{U}$

How to reconstruct the solution?

- After solving the relaxed problem we obtain a matrix $\mathbf{X}^*$
  Q: how to recover the cut?

Goemans-Williamson's algorithm

1. Generate random vector $\mathbf{v}$ in the unit sphere
2. $S = \{i \mid \langle \mathbf{v}, \mathbf{u}_i \rangle \geq 0\}$, where $\mathbf{u}_i$ form matrix $\mathbf{U} : \mathbf{X} = \mathbf{U}^\top \mathbf{U}$

Accuracy estimate

Goemans-Williamson's algorithm in average gives the solution r^* of the original problem, which is not less than

$$c^* \geq r^* \geq 0.878 p^*$$

Authors won the Fulkerson prize in 2000 for this algorithm

Proof

1. Since the method maps any vector $\mathbf{u}_i$ to ± 1 randomly, we can compute the expectation value of the cut C

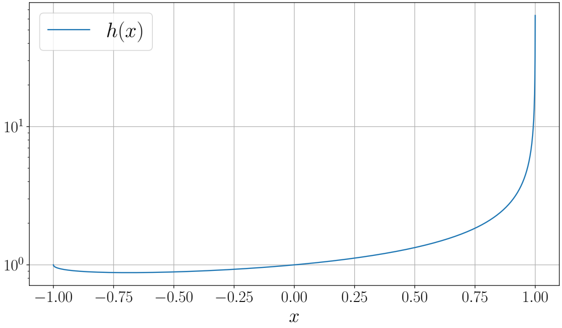
$$\mathbb{E}_{\mathbf{v}}(C) = \mathbb{E}_{\mathbf{v}} \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} [\operatorname{sign}(\mathbf{v}^\top \mathbf{u}_i) \neq \operatorname{sign}(\mathbf{v}^\top \mathbf{u}_j)] \right)$$
$$= \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \mathbb{P}(\operatorname{sign}(\mathbf{v}^\top \mathbf{u}_i) \neq \operatorname{sign}(\mathbf{v}^\top \mathbf{u}_j)) \right)$$

2. Probability of the different signs of $\mathbf{v}^\top \mathbf{u}_i$ and $\mathbf{v}^\top \mathbf{u}_j$ is equal to $\frac{\angle(\mathbf{u}_i, \mathbf{u}_j)}{\pi} = \frac{\arccos(\mathbf{u}_i^\top \mathbf{u}_j)}{\pi}$
3. As a result $\mathbb{E}_{\mathbf{v}}(C) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{\arccos(\mathbf{u}_i^\top \mathbf{u}_j)}{\pi}$

4. Now reduce this expression to the known one

$$\mathbb{E}_{\mathbf{v}}(C) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{2\arccos(\mathbf{u}_i^\top \mathbf{u}_j)}{\pi(1-\mathbf{u}_i^\top \mathbf{u}_j)}\frac{1-\mathbf{u}_i^\top \mathbf{u}_j}{2}$$

5. Find minimum of $h(x) = \frac{2\arccos(x)}{\pi(1-x)}$ for $x \in [-1, 1)$



6. $h(x^*) \approx 0.8785 = \alpha_{GW}$

7. Then

$$\mathbb{E}_{\mathbf{v}}(C) \geq \alpha_{GW} \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(1 - \mathbf{u}_i^\top \mathbf{u}_j) = \alpha_{GW} p^* \geq \alpha_{GW} c^*$$

8. Since Goemans-Williamson's method gives some cut, then

$$c^* \geq \mathbb{E}_{\mathbf{v}}(C) \geq \alpha_{GW} p^* \geq \alpha_{GW} c^*$$

9. Finally

$$c^* \leq p^* \leq \frac{1}{\alpha_{GW}} c^* \approx 1.1382 c^*$$

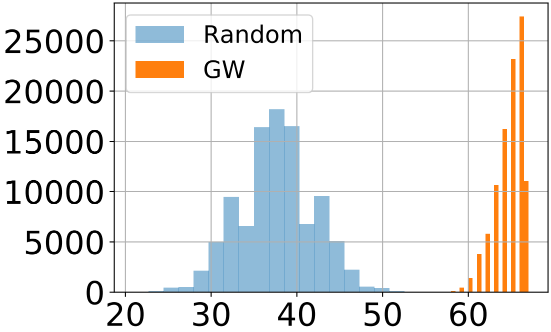
Another interpretation of MAXCUT problem

$$
\begin{aligned}
\max_{x_i=\pm 1} \frac{1}{4}\sum_{i,j=1}^{n} w_{ij}(1-x_ix_j) &= \max_{x_i=\pm 1} \frac{1}{4}\sum_{i,j=1}^{n} w_{ij}\left(\frac{x_i^2+x_j^2}{2} - x_ix_j\right) \\
&= \max_{x_i=\pm 1} \frac{1}{4}\left(-\sum_{i,j=1}^{n} w_{ij}x_ix_j + \frac{1}{2}\sum_{i=1}^{n}\left[\sum_{j=1}^{n} w_{ij}\right] x_i^2 + \frac{1}{2}\sum_{j=1}^{n}\left[\sum_{i=1}^{n} w_{ij}\right] x_j^2\right) \\
&= \max_{x_i=\pm 1} \frac{1}{4}\left(-\sum_{i,j=1}^{n} w_{ij}x_ix_j + \frac{1}{2}\sum_{i=1}^{n}\deg(i)x_i^2 + \frac{1}{2}\sum_{j=1}^{n}\deg(j)x_j^2\right) \\
&= \max_{x_i=\pm 1} \frac{1}{4}\left(\sum_{i=1}^{n}\deg(i)x_i^2 - \sum_{i,j=1}^{n} w_{ij}x_ix_j\right) = \max_{x_i=\pm 1}\frac{1}{4}\mathbf{x}^\top \mathbf{L}\mathbf{x},
\end{aligned}
$$

where $\mathbf{L}$ is a graph Laplacian, $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is a diagonal matrix, where the vertices degrees in the diagonal.

Comparison with random sampling

$$\max_{x_i = \pm 1} \frac{1}{4} \mathbf{x}^\top \mathbf{L} \mathbf{x}$$



Can this bound be improved?

- It is known that the method to compute the approximation with accuracy $\frac{16}{17}$ of the optimal value is already an NP-hard problem!
- Open problem: how to compute the better approximation with the algorithm of subexponential complexity

Summary

- Gradient-free methods
- SDP
- Convex relaxation of non-convex problems
- Goemans-Williamson's algorithm