Gradient Search in the Space of Permutations

An Application for the Linear Ordering Problem

Valentino Santucci, Josu Ceberio, Marco Baioletti

Gradient Search

What is it?

A classical technique for optimizing continuous and differentiable functions.

High popularity due to its application in Neural Network training…

Gradient Search

How it Works?

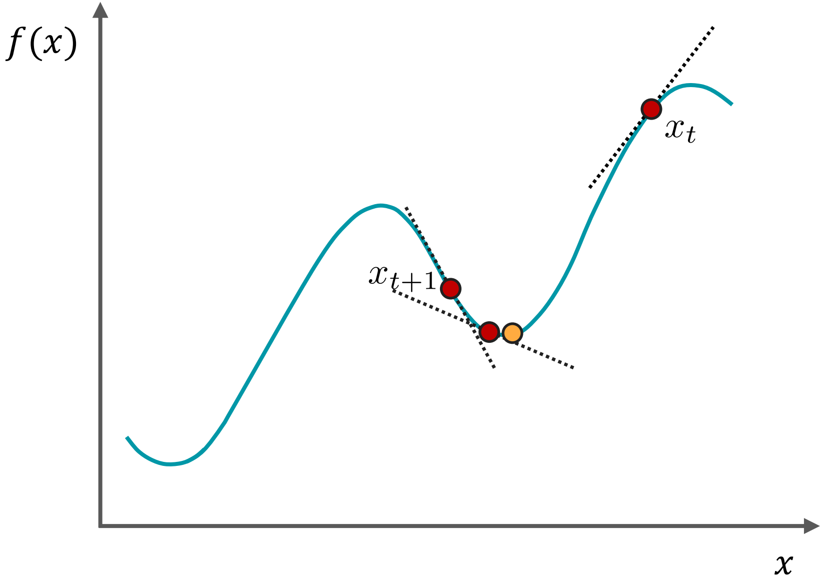
Key Elements:

- **Loss function** (differentiable) to minimize
- Starting point, x_t
- The **gradient** indicates the **direction and magnitude** towards adjust the parameter

Gradient of f at point x_t

$$x_{t+1} \leftarrow x_t + \eta \nabla f(x_t)$$

Learning rate



Gradient Search

For combinatorial optimization?

Combinatorial problems do not have differentiable objective functions!

Objective function

$$F(\theta) = E_\theta[f(x)] = \sum_{x \in X} f(x) p_\theta(x)$$

Probability mass function differentiable w.r.t. θ

A solution:

1. Define **a random variable** that follows a model with **continuous parameters** over the combinatorial space

2. **Optimize the expected value** of the objective function

Gradient Search

For combinatorial optimization?

Objective function

$$F(\theta) = E_\theta[f(x)] = \sum_{x \in X} f(x) p_\theta(x)$$

Probability mass function differentiable w.r.t. θ

Compute the gradients over the parameters of the model!

$$\begin{aligned} \nabla_\theta F(\theta) &= \nabla_\theta \textstyle\sum_{x \in X} f(x) p_\theta(x) \\ &= \ldots \\ &= E_\theta \left[f(x) \nabla_\theta \log p_\theta(x) \right] \end{aligned}$$

Non–affordable, prohibitive size

A solution:

1. Define **a random variable** that follows a model with **continuous parameters** over the combinatorial space

2. **Optimize the expected value** of the objective function

Gradient Search

For combinatorial optimization?

Compute the gradients over the parameters of the model!

$$\begin{aligned}\nabla_\theta F(\theta) &= \nabla_\theta \textstyle\sum_{x \in X} f(x) p_\theta(x) \\ &= \dots \\ &= E_\theta\left[f(x) \nabla_\theta \log p_\theta(x)\right]\end{aligned}$$

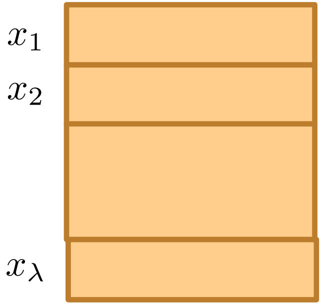
Objective function

$$F(\theta) = E_\theta[f(x)] = \sum_{x \in X} f(x) p_\theta(x)$$

Probability mass function differentiable w.r.t. θ

Non-affordable, prohibitive size

$p_\theta \xrightarrow{\lambda \text{ samples}}$



$\xrightarrow{\text{approx.}}$

$$\nabla_\theta F(\theta) \approx \frac{1}{\lambda} \sum_{i=1}^{\lambda} f(x_i) \nabla_\theta \log p_\theta(x_i)$$

The Space of Permutations

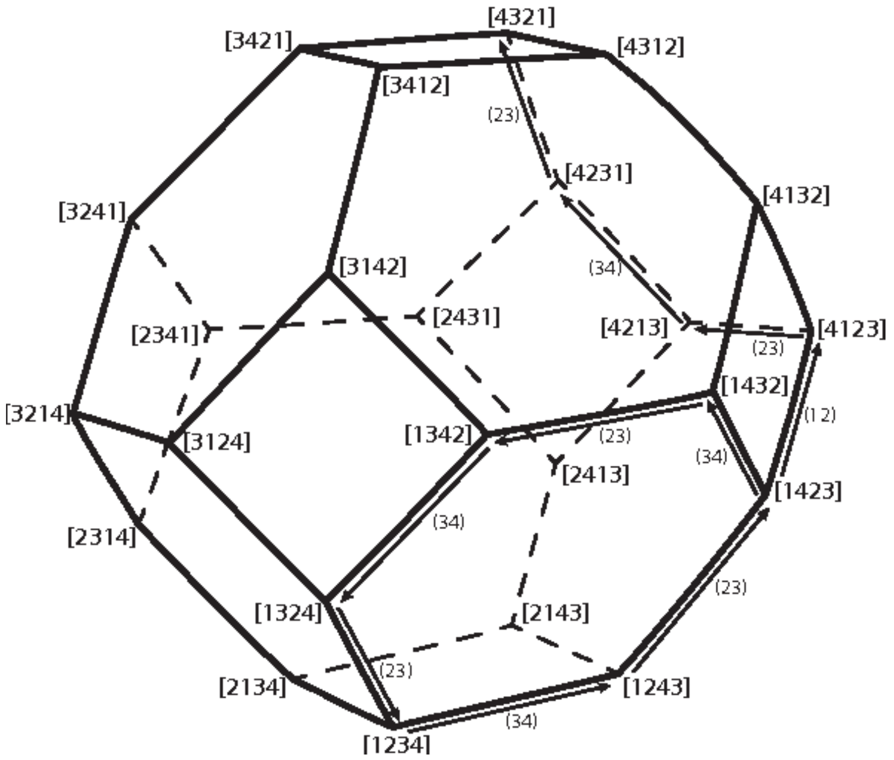
Combinatorial optimization

Algebraic group → SYMMETRIC GROUP

Search space for many problems

Challenges

- Factorial size
- Mutual exclusivity constraint
- Dependencies between items



The 'Butterfly effect' in Cayley graphs, and its relevance for evolutionary genomics.
Vincent Moulton, Mike A. Steel. 2011.

Linear Ordering Problem

A case of study

0	16	11	15	7
21	0	14	15	9
26	23	0	26	12
22	22	11	0	13
30	28	25	24	0

$$f(\sigma) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b_{\sigma(i),\sigma(j)}$$

Linear Ordering Problem

A case of study

	5	3	4	2	1
5	0	25	24	28	30
3	12	0	26	23	26
4	13	11	0	22	22
2	9	14	15	0	21
1	7	11	15	16	0

$$f(\sigma) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b_{\sigma(i),\sigma(j)}$$

Probability Models for Permutations

The Bridge Between Domains

Order Statistic Models

- Babington-Smith
- Bradley-Terry
- Thurstone models
- Plackett-Luce

✓

Distance-Based Models

- Mallows
- Generalized Mallows
- …

✗

Probability Models for Permutations

The Bridge Between Domains

Order Statistic Models

- Babington-Smith
- Bradley-Terry
- Thurstone models
- Plackett-Luce

✓

Distance-Based Models

- Mallows
- Generalized Mallows
- …

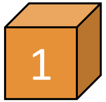
✗

Plackett-Luce Model

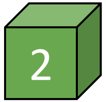
Parameters in $\mathbb{R}$

$$P(\sigma) = \prod_{i=1}^{n} \left(\frac{w_{\sigma_i}}{\sum_{j=i}^{n} w_{\sigma_j}} \right)$$

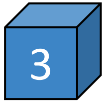
- Each item in the permutation has a weight associated.
- The weights sum up 1.
- The weight associated to each item represents its probability to appear at first rank.



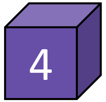
$$w_1 = 0.3$$



$$w_2 = 0.03$$

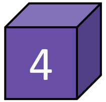


$$w_3 = 0.17$$



$$w_4 = 0.6$$

$$P(\#i) = \frac{w_i}{w_1 + w_2 + w_3 + w_4}$$

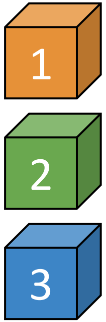


Plackett-Luce Model

Parameters in $\mathbb{R}$

$$P(\sigma) = \prod_{i=1}^{n} \left(\frac{w_{\sigma_i}}{\sum_{j=i}^{n} w_{\sigma_j}} \right)$$

- Each item in the permutation has a weight associated.
- The weights sum up 1.
- The weight associated to each item represents its probability to appear at first rank.

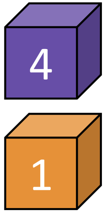


$w_1 = 0.3$

$w_2 = 0.03$

$w_3 = 0.17$

$$P(\#i) = \frac{w_i}{w_1+w_2+w_3}$$

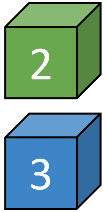


Plackett-Luce Model

Parameters in $\mathbb{R}$

$$P(\sigma) = \prod_{i=1}^{n} \left(\frac{w_{\sigma_i}}{\sum_{j=i}^{n} w_{\sigma_j}} \right)$$

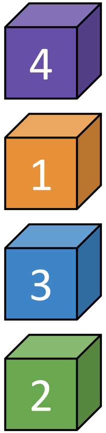
- Each item in the permutation has a weight associated.
- The weights sum up 1.
- The weight associated to each item represents its probability to appear at first rank.



$w_2 = 0.03$

$w_3 = 0.17$

$$P(\#i) = \frac{w_i}{w_2 + w_3}$$

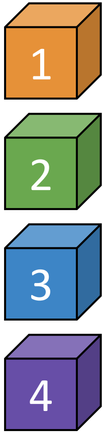


Plackett-Luce Model

Parameters in $\mathbb{R}$

$$P(\sigma) = \prod_{i=1}^{n} \left(\frac{w_{\sigma_i}}{\sum_{j=i}^{n} w_{\sigma_j}} \right)$$

- Each item in the permutation has a weight associated.
- The weights sum up 1.
- The weight associated to each item represents its probability to appear at first rank.

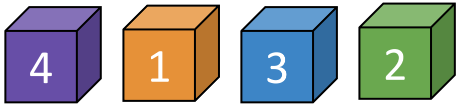


$$w_1 = 0.3$$

$$w_2 = 0.03$$

$$w_3 = 0.17$$

$$w_4 = 0.6$$



$$P(\sigma) = \frac{w_4}{w_1 + w_2 + w_3 + w_4} \cdot \frac{w_1}{w_1 + w_2 + w_3} \cdot \frac{w_3}{w_2 + w_3} \cdot \frac{w_2}{w_2}$$

The Proposal

Gradient search in the space of permutations

Advantages of the PL model:

- Continuous parameters
- Differentiable probability function
- Unbiased sampling procedure

$$P(\sigma|\tilde{\boldsymbol{w}}) = \prod_{i=1}^{n-1} \frac{\exp\left(\tilde{w}_{\sigma(i)}\right)}{\sum_{j=i}^{n} \exp\left(\tilde{w}_{\sigma(j)}\right)}$$

Soft restart of $\tilde{w}$ when convergence

Feasibility Trick!
$\tilde{w} = \log w$

Approximated gradient

$$\nabla_{\tilde{w}} F(\tilde{w}) \approx \frac{1}{\lambda} \sum_{i=1}^{\lambda} U(f(\sigma_i)) \nabla_{\tilde{w}} \log P(\sigma_i|\tilde{w})$$

Utility functions

- Fitness
- Normalized Fitness
- Super linear

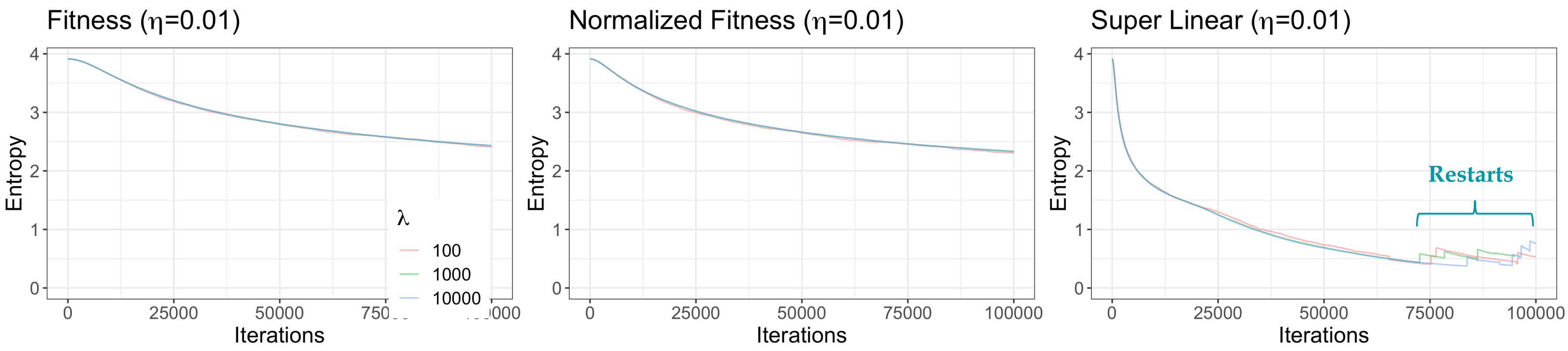
Experimental Study

Impact of utility functions

Experimental design

- $\eta = 0.01$
- $\lambda = \{10^2, 10^3, 10^4\}$
- Stop after 10^5 iterations
- Entropy of the vector of weights

$$H(\mathbf{w}) = -\sum_{i=1}^{n} p_i \log p_i, \qquad p_i = \frac{w_i}{\sum_{j=1}^{n} w_j}$$



λ is not affecting convergence

Preferred convergence

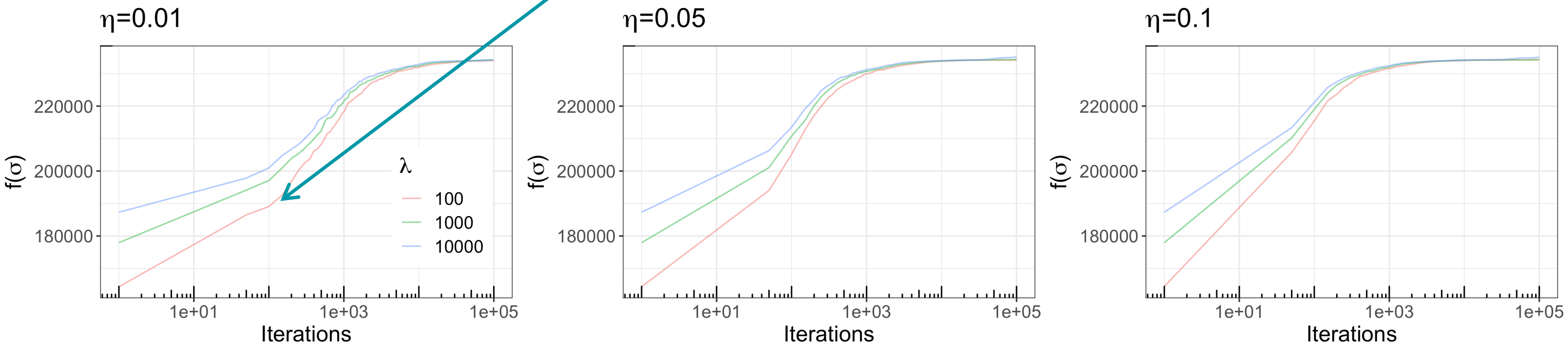
Experimental Study

Tuning η and λ parameters

Experimental design

- $\eta = \{0.01, 0.05, 0.1\}$
- $\lambda = \{10^2, 10^3, 10^4\}$
- Stop after 10^5 iterations
- Best average of 10 repetitions

The smaller λ, the better
Slower convergence with small η



Experimental Study

Comparison to Plackett-Luce EDA

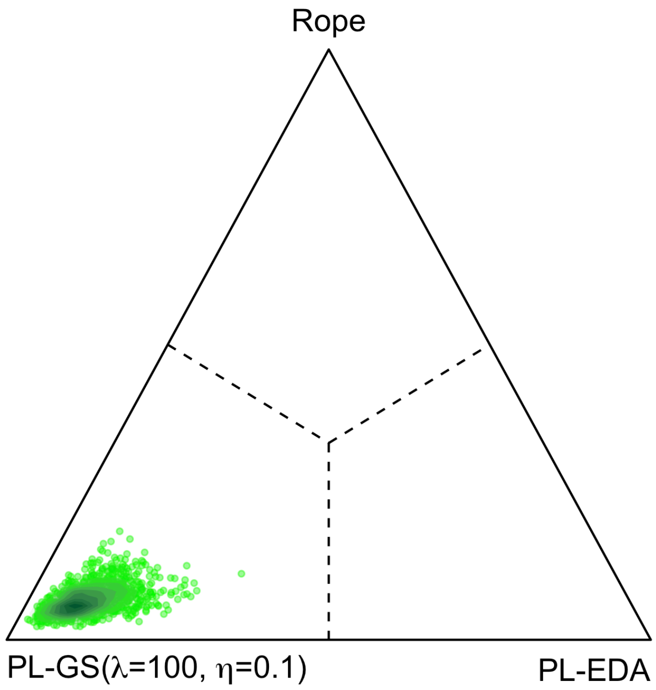
Experimental design

- 20 Repetitions
- Plackett-Luce EDA[1]
- LOLIB benchmark (50 instances)
- $1000n^2$evaluations
- Median Relative Deviation

Gradient Search **better in 40 instances out of 50**

Bayesian statistical analysis [2] - Expected probabilites

- **Gradient Search**: 0.835
- **PL-EDA**: 0.097
- Rope: 0.066



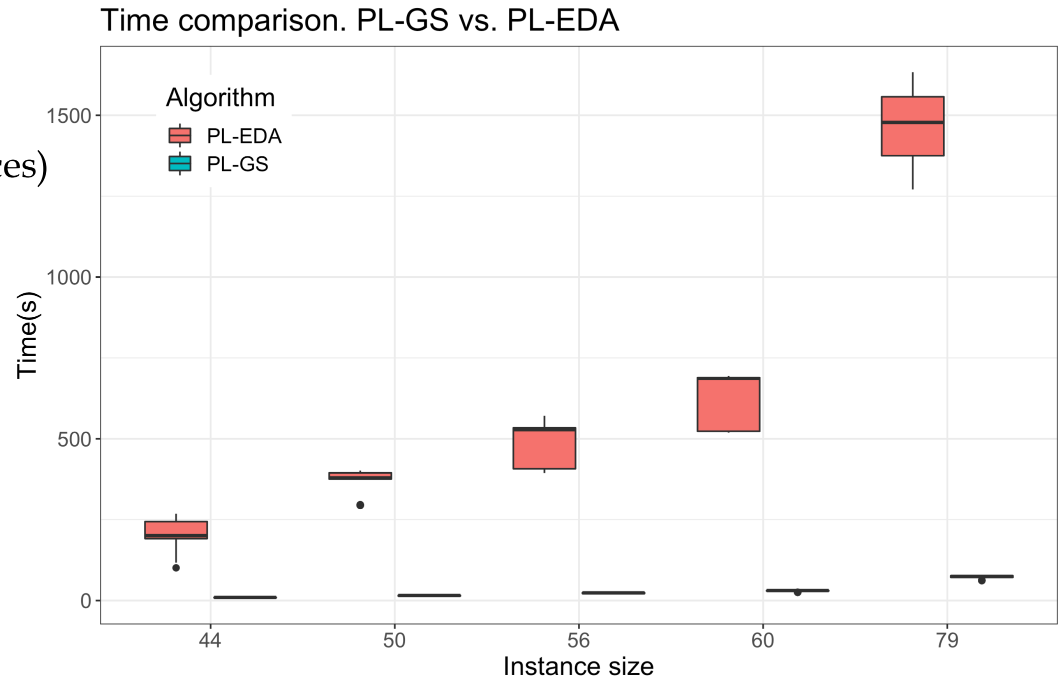
[1] The Plackett-Luce Estimation of Distribution Algorithm. J. Ceberio, A. Mendiburu, J.A. Lozano, 2013.
[2] This statistical analysis is available in the development version of the **scmamp** R package available at **https://github.com/b0rxa/scmamp**.

Experimental Study

Comparison to Plackett-Luce EDA

Experimental design

- 20 Repetitions
- Plackett-Luce EDA[1]
- LOLIB benchmark (50 instances)
- $1000n^2$ evaluations
- Computation time (s)



[1] The Plackett-Luce Estimation of Distribution Algorithm. J. Ceberio, A. Mendiburu, J.A. Lozano, 2013.

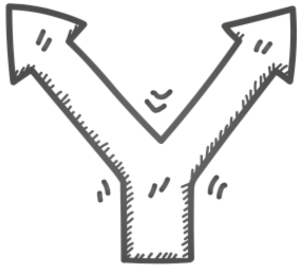
Conclusion & Future Lines

Conclusions

- Devised the mathematical tools for using Gradient Search on the Space of Permutations
- Appropiatedly tuned, promissing performance
- Good computational time

Future lines

- Is the Plackett-Luce **coherent** to any permutation problem? Bradley-Terry…
- Other utility functions?
- Self-adapting η and λ parameters?



Gradient Search in the Space of Permutations

THANK YOU VERY MUCH FOR YOUR ATTENTION!

Valentino Santucci, Josu Ceberio, Marco Baioletti