

# Matrix scaling

Axel Böhm

October 3, 2021

1 Introduction

2 Matchings

3 Permanent

# Introduction

**given:** a matrix  $A \in \mathbb{R}_+^{m \times n}$ , vectors  $r \in \mathbb{R}_{++}^m$  and  $c \in \mathbb{R}_{++}^n$

**find:** diagonal matrices  $X$  and  $Y$  such that for  $B = XAY$  it holds:

$$B\mathbb{1}_n = r \quad \text{and} \quad B^T\mathbb{1}_m = c$$

where  $\mathbb{1}_n = (1, \dots, 1)$  exactly  $n$ -times. Equivalently

$$\|B_{i,:}\|_1 = r_i \quad \text{and} \quad \|B_{:,j}\| = c_j.$$

In this case  $A$  is called  $(r, c)$ -scalable.

If  $\|r\|_1 \neq \|c\|_2$  this is not possible.

# Introduction

**given:** a matrix  $A \in \mathbb{R}_+^{m \times n}$ , vectors  $r \in \mathbb{R}_{++}^m$  and  $c \in \mathbb{R}_{++}^n$

**find:** diagonal matrices  $X$  and  $Y$  such that for  $B = XAY$  it holds:

$$B\mathbb{1}_n = r \quad \text{and} \quad B^T\mathbb{1}_m = c$$

where  $\mathbb{1}_n = (1, \dots, 1)$  exactly  $n$ -times. Equivalently

$$\|B_{i,:}\|_1 = r_i \quad \text{and} \quad \|B_{:,j}\| = c_j.$$

In this case  $A$  is called  $(r, c)$ -scalable.

If  $\|r\|_1 \neq \|c\|_2$  this is not possible.

# Visualization of diagonal scaling

$$B = \begin{bmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_m \end{bmatrix} A \begin{bmatrix} y_1 & & & \\ & y_2 & & \\ & & \ddots & \\ & & & y_n \end{bmatrix}$$

$$= \begin{bmatrix} a_{1,1}x_1y_1 & a_{1,2}x_1y_2 & \cdots & a_{1,n}x_1y_n \\ \vdots & & \ddots & \\ a_{m,1}x_my_1 & & \cdots & a_{m,n}x_my_n \end{bmatrix}$$

**Application:** Ill conditioned linear system  $Az = b$ .

Can multiply both sides by  $X$  and substitute  $z = Yv$  to get instead

$$XAz = X$$

# $(0 - 1)$ matrices | bipartite graphs

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

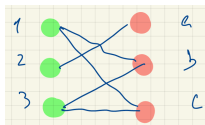


Figure: bipartite graph

## Definition

A **matching** is a set of edges without common vertices.

## Definition

A **perfect matching** is a matching which covers all vertices.

# $(0 - 1)$ matrices | bipartite graphs

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

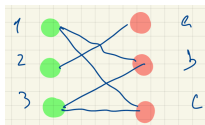


Figure: bipartite graph

## Definition

A **matching** is a set of edges without common vertices.

## Definition

A **perfect matching** is a matching which covers all vertices.

# Finding the number of perfect matchings

Finding one is easy (polynomial time). Finding all is in  $\# P$  (i.e. hard!).

Consider  $m = n$ ,  $A \in \mathbb{R}^{n \times n}$

Recall:

$$\text{(determinant)} \quad \det A = \sum_{\sigma} \text{sign}(\sigma) \prod_{i=1}^n a_{i\sigma(i)}$$

$$\text{(permanent)} \quad \text{perm } A = \sum_{\sigma} \prod_{i=1}^n a_{i\sigma(i)}$$

## Observation

For a  $(0, 1)$ -matrix  $A$ ,  $\text{perm } A$  is the number of perfect matchings.

One is easy to compute the other one hard. How can this be?



# Lower bounding the permanent

## Definition

A matrix  $A \in \mathbb{R}_+^{m \times n}$  is called **doubly stochastic**, if sum of every row and every column is 1.

van der Waerden (1926) conjectured

For doubly stochastic matrices the following *lower bound* holds

$$\text{perm } A \geq \frac{n!}{n^n}.$$

Is tight for  $A = \begin{bmatrix} 1/n & \dots & 1/n \\ \vdots & \ddots & 1/n \\ 1/n & \dots & 1/n \end{bmatrix}$

Proved independently by Jegorščow and Falikman in '80 / '81.

# Upper bounding the permanent

## Bregman-Minc

For  $(0, 1)$ -matrices

$$\text{perm } A \leq \prod_{i=1}^n (r_i!)^{1/r_i} \quad \text{where } r_i := \|A_{i,:}\|$$

# Matrix scaling to approx. permanent

If a  $(0, 1)$ -matrix  $A$  can be scaled to be doubly stochastic, i.e. it is  $(\mathbb{1}, \mathbb{1})$ -scalable, then we can apply lower bound

$$\text{perm } B = \text{perm}(XAY) = \left( \prod_i x_i \right) \left( \prod_j y_j \right) \text{perm} A$$

# Matrix scaling as an optimization problem

- ◇ **given:**  $A, r, c$
- ◇ **find:**  $X, Y$  such that  $B = XAY$  fulfills  $B\mathbb{1}_m = r$  and  $B\mathbb{1}_n = c$ .
- ◇  $m + n$  unknowns
- ◇  $m + n$  constraints

Consider the (*nonconvex*) function

$$g(x, y) = \langle x, Ay \rangle - \langle r, \log x \rangle - \langle c, \log y \rangle$$

with derivative (coordinatewise)

$$\begin{aligned}\nabla_x g(x, y) &= Ay - \frac{r}{x} \\ \nabla_y g(x, y) &= A^T x - \frac{c}{y}\end{aligned}\tag{1}$$

# Reparametrizing this system

Via reparametrization  $x = e^\xi$  and  $y = e^\eta$  we get

$$f(\xi, \eta) = \sum_{i,j} a_{i,j} e^{\xi_i + \eta_j} - \langle r, \xi \rangle - \langle c, \eta \rangle$$

which is *convex*. It's gradient is given by

$$\frac{\partial f}{\partial \xi_i} = \sum_{j=1}^n a_{i,j} e^{\xi_i + \eta_j} - r_i \quad (2)$$

Easy to see that the optimality condition of (2) and (1) agree.  
Implies that even the nonconvex function only has *global* minimizers.

# Matrix scaling as an optimization problem [contd]

It is easy to see that a solution  $(x, y)$  of

$$\begin{aligned} Ay - \frac{r}{x} &= 0 \\ A^T x - \frac{c}{y} &= 0 \end{aligned}$$

defines a solution to the *matrix scaling* problem via  $X = \text{diag } x$  and  $Y = \text{diag } y$

$$\begin{pmatrix} a_{11}y_1 + a_{12}y_2 + \cdots \\ a_{21}y_1 + a_{22}y_2 + \cdots \\ a_{m1}y_1 + a_{m2}y_2 + \cdots \end{pmatrix} \cdot x_i = r_i$$

The question remains: how to minimize

$$g(x, y) = \langle x, Ay \rangle - \langle r, \log x \rangle - \langle c, \log y \rangle$$

alternating minimization

Given a problem

$$\min_{x, y} \varphi(x, y)$$

$$x_{k+1} = \arg \min_x \varphi(x, y_k)$$

$$y_{k+1} = \arg \min_y \varphi(x_{k+1}, y)$$

makes sense as long as the subproblems are easy (e.g. convex).

$$g(x, y) = \langle x, Ay \rangle - \langle r, \log x \rangle - \langle c, \log y \rangle$$

$$\text{opt. cond. for } x \quad Ay - \frac{r}{x} = 0 \quad \text{opt. cond. for } y \quad Ay - \frac{c}{y} = 0$$

The question remains: how to minimize

$$g(x, y) = \langle x, Ay \rangle - \langle r, \log x \rangle - \langle c, \log y \rangle$$

### alternating minimization

Given a problem

$$\min_{x, y} \varphi(x, y)$$

$$x_{k+1} = \arg \min_x \varphi(x, y_k)$$

$$y_{k+1} = \arg \min_y \varphi(x_{k+1}, y)$$

makes sense as long as the subproblems are easy (e.g. convex).

$$g(x, y) = \langle x, Ay \rangle - \langle r, \log x \rangle - \langle c, \log y \rangle$$

$$\text{opt. cond. for } x \quad Ay - \frac{r}{x} = 0 \quad \text{opt. cond. for } y \quad Ay - \frac{c}{y} = 0$$



## Sinkhorn '60

Given  $(x_0, y_0)$ , for  $k = 1, \dots$

$$x_{k+1} = \frac{r}{Ay_k}$$
$$y_{k+1} = \frac{r}{Ax_{k+1}}$$

Linear convergence if  $a_{i,j} > 0$ . **Q:** What if  $A$  is not  $(r, c)$ -scalable?