Matrix scaling

Axel Böhm

November 11, 2021

Introduction

given: a matrix $A \in \mathbb{R}^{m \times n}_{\geq 0}$, vectors $r \in \mathbb{R}^m_{>0}$ and $c \in \mathbb{R}^n_{>0}$
find: nonneg. diagonal matrices X and Y such that for

$$B = XAY$$

it holds that:

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

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

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

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

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

Introduction

given: a matrix $A \in \mathbb{R}_{\geq 0}^{m \times n}$, vectors $r \in \mathbb{R}_{>0}^{m}$ and $c \in \mathbb{R}_{>0}^{n}$
find: nonneg. diagonal matrices X and Y such that for

$$B = XAY$$

it holds that:

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

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

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

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

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

Visualization of diagonal scaling

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

$$= \begin{bmatrix} a_{1,1}x_1y_1 & a_{1,2}x_1y_2 & \cdots & a_{1,n}x_1y_m \\ \vdots & \ddots & & \\ a_{m,1}x_my_1 & & \cdots & a_{m,n}x_my_m \end{bmatrix}$$

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

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

$$XAYv = Xb$$

$(0-1)$ matrices | bipartite graphs

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

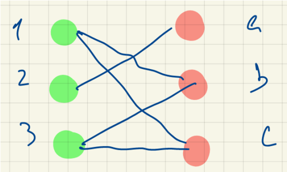

Figure: bipartite graph

Definition

- A **matching** is a set of edges without common vertices.
- A **perfect matching** is a matching which covers all vertices.

Applications:

- marriage problem
- Hitchcock transport problem

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:

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

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

Observation

For a $(0,1)-$matrix A, $\operatorname{perm} A$ 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 $A \in \mathbb{R}_{\geq 0}^{m \times n}$ 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

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

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

Proved in 1980.

Matrix scaling to approximate the 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

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

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

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

Easy to see that the optimality condition of (2) and (1) agree.
$\Rightarrow$ 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

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

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

$$\begin{pmatrix} a_{11}y_1 + a_{12}y_2 + \cdots \\ a_{21}y_1 + a_{22}y_2 + \cdots \\ a_{m1}y_1 + a_{m2}y_2 + \cdots \end{pmatrix} \begin{matrix} \cdot x_1 = r_1 \\ \cdot x_2 = r_2 \\ \cdot x_m = r_m \end{matrix}$$

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

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

repeat:

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

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

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

Sinkhorn's Algorithm

Sinkhorn '60

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

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

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