The solution to $\arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}\|_2^2$ subject to $\langle \mathbf{x}, \mathbf{1} \rangle = 1$ is $\mathbf{x}^* = \dfrac{(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1}}{\mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1}}$

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft: August 14, 2019
Last update : August 14, 2019

A constrained problem

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with rank n, find $\mathbf{x}^*$ by solving

$$(\mathcal{P}) \ : \ \mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^n}{\arg\min} \|\mathbf{A}\mathbf{x}\|_2^2 \ \text{ subjec to } \langle \mathbf{x}, \mathbf{1} \rangle = 1$$

where $\mathbf{1}$ is all-1 vector in $\mathbb{R}^n$.

- The problem means find a vector $\mathbf{x}$ in $\mathbb{R}^n$ such that it minimizes $\|\mathbf{A}\mathbf{x}\|_2^2$ while all its elements sum to 1
- It is possible to replace $\|\mathbf{A}\mathbf{x}\|_2^2$ with $\|\mathbf{A}\mathbf{x}\|_2^1$, the square is just for convince of taking derivatives
- An example where problem $(\mathcal{P})$ appear : Anderson Acceleration

It can be shown that, this problem has analytic close form solution as

$$\mathbf{x}^* = \frac{(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{1}}{\mathbf{1}^\top (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{1}}$$

This document : show how to prove this.

Lagrangian

The problem

$$(\mathcal{P}) \ : \ \mathbf{x}^* = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} \|\mathbf{A}\mathbf{x}\|_2^2 \ \text{ subjec to } \langle \mathbf{x}, \mathbf{1} \rangle = 1$$

is a problem with equality constraint. Hence we solve it by consider the Lagrangian : let λ be the Lagrangian multiplier, we have

$$L(\mathbf{x}, \lambda) = \|\mathbf{A}\mathbf{x}\|_2^2 + \lambda\big(\langle \mathbf{x}, \mathbf{1} \rangle - 1\big)$$

The solution of $(\mathcal{P})$ can be found by solving the following system of equations

$$\begin{aligned} \frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} &= 0 \\ \frac{\partial L(\mathbf{x}, \lambda)}{\partial \lambda} &= 0 \end{aligned}$$

On details of minimizing the Lagrangian ... (1/3)

For $L(\mathbf{x}, \lambda) = \|\mathbf{A}\mathbf{x}\|_2^2 + \lambda\big(\langle \mathbf{x}, \mathbf{1}\rangle - 1\big)$, we have

$$
\begin{aligned}
\frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} &= 2\mathbf{A}^\top \mathbf{A}\mathbf{x} + \lambda \mathbf{1} = 0 \\
\frac{\partial L(\mathbf{x}, \lambda)}{\partial \lambda} &= \langle \mathbf{x}, \mathbf{1}\rangle - 1 = 0
\end{aligned}
$$

(Recall, the derivative of $\langle \mathbf{x}, \mathbf{a}\rangle$ w.r.t. $\mathbf{x}$ is $\mathbf{a}$).
Therefore, the optimal pair $(\mathbf{x}^*, \lambda^*)$ fulfil the KKT system :

$$
\begin{bmatrix} 2\mathbf{A}^\top \mathbf{A} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}
\begin{bmatrix} \mathbf{x}^* \\ \lambda^* \end{bmatrix}
=
\begin{bmatrix} 0 \\ 1 \end{bmatrix}
$$

Our goal is to find the expression $\mathbf{x}^*$, hence we solve the KKT system

$$
\begin{bmatrix} \mathbf{x}^* \\ \lambda^* \end{bmatrix}
=
\begin{bmatrix} 2\mathbf{A}^\top \mathbf{A} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}^{-1}
\begin{bmatrix} 0 \\ 1 \end{bmatrix}
$$

The matrix is a 2-by-2 block matrix, we need to apply Schur complement.

On details of minimizing the Lagrangian ... (2/3)

For 2-by-2 block matrix $\mathbf{R} = \begin{bmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}$, if $\mathbf{W}$ is non-singular, then

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{W}^{-1} + \mathbf{W}^{-1}\mathbf{X}(\mathbf{Z} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{Y}\mathbf{W}^{-1} & -\mathbf{W}^{-1}\mathbf{X}(\mathbf{Z} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1} \\ -(\mathbf{Z} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{Y}\mathbf{W}^{-1} & (\mathbf{Z} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1} \end{bmatrix}$$

For $\mathbf{R} = \begin{bmatrix} 2\mathbf{A}^\top\mathbf{A} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}$ we have $\mathbf{Z} = 0$, then

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{X}(\mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{Y}\mathbf{W}^{-1} & \mathbf{W}^{-1}\mathbf{X}(\mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1} \\ (\mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{Y}\mathbf{W}^{-1} & -(\mathbf{Y}\mathbf{W}^{-1}\mathbf{X})^{-1} \end{bmatrix}$$

As $\mathbf{X} = \mathbf{1}$, $\mathbf{Y} = \mathbf{1}^\top$, $\mathbf{W} = 2\mathbf{A}^\top\mathbf{A}$, we have $\mathbf{Y}\mathbf{W}^{-1}\mathbf{X} = \frac{1}{2}\mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1}$, which is a scalar, denote it as a, we have

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{X}a^{-1}\mathbf{Y}\mathbf{W}^{-1} & \mathbf{W}^{-1}\mathbf{X}a^{-1} \\ a^{-1}\mathbf{Y}\mathbf{W}^{-1} & -a^{-1} \end{bmatrix}$$

Factorize the a out

$$\begin{aligned} \mathbf{R}^{-1} &= \frac{1}{a}\begin{bmatrix} a\mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{X}\mathbf{Y}\mathbf{W}^{-1} & \mathbf{W}^{-1}\mathbf{X} \\ \mathbf{Y}\mathbf{W}^{-1} & -1 \end{bmatrix} \\ &= \frac{1}{a}\begin{bmatrix} \mathbf{W}^{-1}(a\mathbf{I} - \mathbf{X}\mathbf{Y}\mathbf{W}^{-1}) & \mathbf{W}^{-1}\mathbf{X} \\ \mathbf{Y}\mathbf{W}^{-1} & -1 \end{bmatrix} \end{aligned}$$

Put $\mathbf{X} = \mathbf{1}$, $\mathbf{Y} = \mathbf{1}^\top$ and $\mathbf{W} = 2\mathbf{A}^\top\mathbf{A}$

$$\begin{aligned}
\mathbf{R}^{-1} &= \frac{1}{a}\begin{bmatrix} (2\mathbf{A}^\top\mathbf{A})^{-1}\big(a\mathbf{I} - \mathbf{1}\mathbf{1}^\top(2\mathbf{A}^\top\mathbf{A})^{-1}\big) & (2\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1} \\ \mathbf{1}^\top(2\mathbf{A}^\top\mathbf{A})^{-1} & -1 \end{bmatrix} \\
&= \frac{2}{\mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1}}\begin{bmatrix} \frac{1}{2}(\mathbf{A}^\top\mathbf{A})^{-1}\big(a\mathbf{I} - \mathbf{1}\mathbf{1}^\top\frac{1}{2}(\mathbf{A}^\top\mathbf{A})^{-1}\big) & \frac{1}{2}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1} \\ \mathbf{1}^\top\frac{1}{2}(\mathbf{A}^\top\mathbf{A})^{-1} & -1 \end{bmatrix} \\
&= \frac{1}{\mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1}}\begin{bmatrix} (\mathbf{A}^\top\mathbf{A})^{-1}\big(a\mathbf{I} - \mathbf{1}\mathbf{1}^\top\frac{1}{2}(\mathbf{A}^\top\mathbf{A})^{-1}\big) & (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1} \\ \mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1} & -2 \end{bmatrix} \\
&= \frac{1}{\mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1}}\begin{bmatrix} \frac{1}{2}(\mathbf{A}^\top\mathbf{A})^{-1}\Big(\mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1}\,\mathbf{I} - \mathbf{1}\mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1}\Big) & (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1} \\ \mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1} & -2 \end{bmatrix}
\end{aligned}$$

With this $\mathbf{R}^{-1}$, we can now compute the optimal $\mathbf{x}^*$, which gives

$$\begin{bmatrix} \mathbf{x}^* \\ \lambda^* \end{bmatrix} = \mathbf{R}^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\mathbf{1}^\top(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1}}\begin{bmatrix} (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{1} \\ -2 \end{bmatrix} \quad \square$$

End of document