LINEAR ALGEBRA REVIEW

A PREPRINT

Kai Yi*
Department of Computer Science
King Abdullah University of Science and Technology
kai.yi@kaust.edu.sa

April 29, 2020

1 Linear Algebra Review

This review is inspired from Prof. Zico Kolter [1] and Prof. Chuong Do's modifications at [2].

1.1 Operations and Properties of Matrices

Identity matrix $I \in \mathbb{R}^{n\times n}, I_{ij} = 1 \text{ if } i = j \text{ else } 0.$
Properties: $A \in \mathbb{R}^{m\times n}$, $AI_1 = A = I_2A$, where $I_1 \in \mathbb{R}^{n\times n}, I_2 \in \mathbb{R}^{m\times m}$.

Diagonal Matrix $D = \text{diag}(d_1, \cdots, d_n) \in \mathbb{R}^{n\times n}, D_{ij} = d_i \text{ if } i = j \text{ else } 0.$
Obviously, $I = \text{diag}(1, \cdots, 1)$.

Transpose $A \in \mathbb{R}^{m\times n}$, $(A^T)_{ij} = A_{ji}$.
Properties: $\alpha : (A^T)^T = A;\ \ \beta : (AB)^T = B^T A^T;\ \ \gamma : (A + B)^T = A^T + B^T.$

Symmetric Matrices $A \in \mathbb{R}^{n\times n}$ is **symmetric** if $A = A^T$, is **anti-symmetric** if $A = -A^T$.
Obviously, $A + A^T$ is symmetric while $A - A^T$ is anti-symmetric.
Any $A \in \mathbb{R}^{n\times n}$ can be represented as a sum of a symmetric matrix and an anti-symmetric matrix:

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T). \tag{1}$$

Trace $A \in \mathbb{R}^{n\times n}$, $tr(A) = \sum_{i=1}^{n} A_{ii}$.
Properties: Assume $A, B \in \mathbb{R}^{n\times n}, t \in \mathbb{R} \models$

$$a : tr(A) = tr(A^T),\ \ b : tr(A + B) = tr(A) + tr(B),\ \ c : tr(tA) = t\, tr(A).$$

Suppose $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times m} \models tr(AB) = tr(BA)$.

*The author is a MS/PhD candidate at King Abdullah University of Science and Technology (KAUST) under the supervision of Prof. Mohamed Elhoseiny. His current research interests include zero shot learning, continual learning, generative models and Bayesian neural network. His homepage is kaiyi.me.

[2] http://cs229.stanford.edu/section/cs229-linalg.pdf

Proof 1

$$
\begin{aligned}
tr(AB) &= \sum_{i=1}^{m}(AB)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ji} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m} B_{ji}A_{ij} = \sum_{j=1}^{n}(BA)_{jj} = tr(BA).
\end{aligned}
\tag{2}
$$

Generally, suppose $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times k}, C \in \mathbb{R}^{k\times m} \models tr(ABC) = tr(BCA) = tr(CAB)$.

Norms Given $X \in \mathbb{R}^n$, $||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$.

Obviously, we have $||x||_2^2 = x^T x$.

Properties a norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies

non-negativity: $f(x) \geq 0$; definiteness: $f(x) = 0$ iif $x = 0$; homogeneity: $t \in \mathbb{R}, f(tx) = |t|f(x)$.

triangle inequality: $x, y \in \mathbb{R}^n, f(x + y) \leq f(x) + f(y)$

Proof 2 $x, y \in \mathbb{R}^n, f(x + y) \leq f(x) + f(y)$.

For left:

$$
f^2(x+y) = \sum_{i=1}^{n}(x_i + y_i)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 + 2\sum_{i=1}^{n} x_i y_i \tag{3}
$$

For right:

$$
(f(x) + f(y))^2 = f^2(x) + f^2(y) + 2f(x)f(y) = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 + 2\sqrt{\sum_{i=1}^{n} x_i^2 y_i^2} \tag{4}
$$

Subtract Eqn (3) by Eqn (4), we have

$$
2\left(\sum_{i=1}^{n} x_i y_i - \sqrt{\sum_{i=1}^{n} x_i^2 y_i^2}\right) \tag{5}
$$

Using The Cauchy-Schwarz Inequality (For $x, y \in \mathbb{R}^n$, $|x \cdot y| \leq ||x||\ ||y||$), we can get Eqn 5 ≤ 0. Thus we have $f^2(x + y) \leq (f(x) + f(y))^2$.

For $f(x + y), f(x), f(y) \geq 0$, we have $f(x + y) \leq f(x) + f(y)$.

ℓ_p **norms** $p \in \mathbb{R}$ and $p \geq 1$,

$$
\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}. \tag{6}
$$

Particularly, $\ell_1 = \|x\|_1 = \sum_{i=1}^{n} |x_i|$, $\ell_\infty = \|x\|_\infty = \max_i |x_i|$.

When we define norms for matrices, for the ℓ_2 norm converted to Frobenius norm,

$$
\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{\operatorname{tr}(A^T A)}. \tag{7}
$$

Linearly dependent/independent If $x_1, \cdots, x_n \subset \mathbb{R}^m, x_n = \sum_{i=1}^{n-1} \alpha_i x_i$ for some scalar values $\alpha_1, \cdots, \alpha_{n-1} \in \mathbb{R}$, then we say that thee vectors $x_1, \cdots, x_n$ are linearly dependent; otherwise, the vectors are linearly independent.

Rank of matrices $A \in \mathbb{R}^{m\times n}$,the size of the largest subset of columns of A that constitute a linearly independent set is **column rank**; similarly, we get **row rank**.

For any matrix $A \in \mathbb{R}^{m\times n}$, column rank equals to row rank. They are collectively called the **rank** of A, denoted as $rank(A)$.

Properties For $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times k}$

$a : rank(A) \leq \min(m, n)$, when $rank(A) = \min(m, n)$, then A is said to be **full rank**.

$b : rank(A) = rank(A^T),\ \ c : rank(AB) \leq \min(rank(A), rank(B)),\ \ rank(A+B) \leq rank(A)+rank(B)$.

Inverse If $A \in \mathbb{R}^{n\times n}$, $A^{-1}A = I = AA^{-1}$, A^{-1} is denoted as the inverse matrix.

We say A is **invertible** or **non-singular** if A^{-1} exists and **non-invertible** or **singular** otherwise.

If A is *invertible*, then A must be full rank.

Properties For $A, B \in \mathbb{R}^{n\times n}$ are non-singular,

$a : (A^{-1})^{-1} = A,\ \ b : (AB)^{-1} = B^{-1}A^{-1},\ \ c : (A^{-1})^T = (A^T)^{-1}$, denoted as A^{-T}.

Orthogonal Matrices A square matrix $U \in \mathbb{R}^{n\times n}$ is **orthogonal** if all its columns are orthogonal to each other and are normalized.

For vectors $x, y \in \mathbb{R}^n$ are **orthogonal** if $x^T y = 0$. A vector $x \in \mathbb{R}^n$ is **normalized** if $||x||_2 = 1$.

For orthogonal matricx, we have

$$U^T U = I = UU^T. \tag{8}$$

Properties $U \in \mathbb{R}^{n\times n}, x \in \mathbb{R}^n$, we have $||Ux||_2 = ||x||_2$.

Proof 3

$$||Ux||_2^2 = (Ux)^T(Ux) = x^T U^T U x = x^T x = ||x||_2^2. \tag{9}$$

Span For $\{x_1, \ldots x_n\} \in \mathbb{R}^m$,

$$\text{span}\left(\{x_1, \ldots x_n\}\right) = \left\{v : v = \sum_{i=1}^{n} \alpha_i x_i, \quad \alpha_i \in \mathbb{R}\right\}. \tag{10}$$

If $\{x_1, \ldots x_n\}$ i.i.d, $x_i \in \mathbb{R}^m$, then $\text{span}\left(\{x_1, \ldots x_n\}\right) = \mathbb{R}^m$.

Projection

$$\text{Proj}\left(y; \{x_1, \ldots x_n\}\right) = \text{argmin}_{v\in \text{span}(\{x_1,\ldots,x_n\})} \|y - v\|_2. \tag{11}$$

Range (Columnspace) Suppose $A \in \mathbb{R}^{m\times n}$, denoted $\mathbb{R}(A)$ is the sapce of the columns of A,

$$\mathbb{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}. \tag{12}$$

If A is full rank and $n \leq m$, then

$$\text{Proj}(y; A) = \text{argmin}_{v\in\mathbb{R}(A)} \|v - y\|_2 = A\left(A^T A\right)^{-1} A^T y. \tag{13}$$

If A contains only a single column, $a \in \mathbb{R}^m$, then

$$\text{Proj}(y; a) = \frac{aa^T}{a^T a} y. \tag{14}$$

Nullspace Given $A \in \mathbb{R}^{m \times n}$, denoted $\mathcal{N}(A)$,

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}. \tag{15}$$

As $|\mathbb{R}(A)| = m, |\mathcal{N}(A)| = n$, then $|\mathbb{R}(A^T)| = |\mathcal{N}(A)| = n$, we can get

$$\left\{w : w = u + v, u \in \mathbb{R}\left(A^T\right), v \in \mathcal{N}(A)\right\} = \mathbb{R}^n \text{ and } \mathbb{R}\left(A^T\right) \cap \mathcal{N}(A) = \{0\}. \tag{16}$$

$\mathcal{R}(\mathcal{A}^\mathcal{T})$ and $\mathcal{N}(A)$ are disjoint subsets that together span the entire space of $\mathbb{R}^n$, called **orthogonal complements**, denoted $\mathbb{R}(A^T) = \mathcal{N}(A)^\perp$.

Determinant $A \in \mathbb{R}^{n \times n}$, is a function det: $\mathbb{R}^{n \times n} \to \mathbb{R}$, and is denoted $|A|$ or det$A$.

One definition of determinant is as follows. We first define

$$S = \left\{v \in \mathbb{R}^n : v = \sum_{i=1}^n \alpha_i a_i \text{ where } 0 \le \alpha_i \le 1, i = 1, \ldots, n\right\}. \tag{17}$$

Then the absolute value of the determinant of A, it turns out, is a measure of the 'volumne' of the set S.

Properties For $A, B \in \mathbb{R}^{n \times n}$,

$a : |A| = |A^T|$, $\;b : |AB| = |A||B|$, $\;c : |A| = 0 \; iif . A$ is singular. $d : A$ non-singular, $|A^{-1}| = 1/|A|$.

General Definition of Determinant For $A \in \mathbb{R}^{n \times n}$, $A_{-i,-j} \in \mathbb{R}^{(n-1)\times(n-1)}$

$$|A| = \sum_{i=1}^n (-1)^{i+j} a_{ij} \left|A_{-i,-j}\right| = \sum_{i=1}^n (-1)^{i+j} a_{ij} \left|A_{-i,-j}\right|. \tag{18}$$

The initial case is $|A| = a_{11} \text{for} A \in \mathbb{R}^{1 \times 1}$.

Several common cases:

$$\begin{aligned} \|a_{11}\| &= a_{11} \\ \left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| &= a_{11}a_{22} - a_{12}a_{21} \\ \left|\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right| &= \begin{array}{c} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{array} \end{aligned} \tag{19}$$

Classical adjoint For matrix $A \in \mathbb{R}^{n \times n}$,

$$\text{adj}(A) \in \mathbb{R}^{n \times n}, \quad (\text{adj}(A))_{ij} = (-1)^{i+j} |A_{-j,-i}|. \tag{20}$$

Note the switch indices $A_{-j,-i}$. For any non-singular $A \in \mathbb{R}^{n \times n}$,

$$A^{-1} = \frac{1}{|A|} \text{adj}(A). \tag{21}$$

Quadratic Forms Given $A \in \mathbb{R}^{n \times n}, x\mathbb{R}^n$, the scalar value $x^T Ax$ is called a *quadratic form*. Explicitly,

$$x^T Ax = \sum_{i=1}^n x_i (Ax)_i = \sum_{i=1}^n x_i \left(\sum_{j=1}^n A_{ij} x_j\right) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j. \tag{22}$$

Where also,

$$x^T A x = \left(x^T A x\right)^T = x^T A^T x = x^T \left(\frac{1}{2}A + \frac{1}{2}A^T\right) x. \tag{23}$$

We can get the following definitions:

- A symmetric matrix $A \in \mathrm{S}^n$ is positive definite (PD) if for all non-zero vectors $x \in \mathbb{R}^n, x^T A x > 0$. This is usually denoted $A \succ 0$, and often times the set of all positive definite matrices is denoted $\mathbf{S}^n_{++}$.
- A symmetric matrix $A \in \mathrm{S}^n$ is positive semidefinite (PSD) if for all vectors $x^T A x \geq 0$. This is written $A \succeq 0$, and the set of all positive semidefinite matrices is often denoted S^n_{+}.
- Likewise, a symmetric matrix $A \in \mathrm{S}^n$ is negative definite (ND), denoted $A \prec 0$ if for all non-zero $x \in \mathbb{R}^n, x^T A x < 0$.
- Similarly, a symmetric matrix $A \in \mathrm{S}^n$ is negative semidefinite (NSD), denoted $A \preceq 0$ if for all $x \in \mathbb{R}^n, x^T A x \leq 0$.
- Finally, a symmetric matrix $A \in \mathrm{S}^n$ is indefinite, if it is neither positive semidefinite nor negative semidefinite - i.e., if there exists $x_1, x_2 \in \mathbb{R}^n$ such that $x_1^T A x_1 > 0$ and $x_2^T A x_2 < 0$.

Proof 4 *Positive definite and negative definite matrices are always full rank.*

Suppose some matrix $A \in \mathbb{R}^{n\times n}$ is not full rank.

Then suppose $a_j = \sum_{i\neq j} x_i a_i$, for some $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n \in \mathbb{R}$.

Setting $x_j = -1$, we have $Ax = \sum_{i=1} x_i a_i = 0$.

Gram matrix Given $A \in \mathbb{R}^{m\times n}$, $G = A^T A$ is always positive semidefinite.

Further, if $m \geq n$ and A is full rank, then G is positive definite.

Eigenvalues and Eigenvectors Given $A \in \mathbb{R}^{n\times n}$, we say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A and $x \in \mathbb{C}^n$ is the corresponding *eigenvector* if $Ax = \lambda x,\ x \neq 0$.

We can also rewrite the representation as an eigenvalue-eigenvector pair of A if $(\lambda I - A)x = 0, x \neq 0$.

Due to $(\lambda I - A)x = 0$ has a non-zero solution to x iif. $(\lambda I - A)$ has a non-empty nullspace, which is only the case if $(\lambda I - A)$ is singular, i.e., $|(\lambda I - A)| = 0$.

Properties

- The trace of a A is equal to the sum of its eigenvalues $\operatorname{tr} A = \sum_{i=1}^n \lambda_i$.
- The determinant of A is equal to the product of its eigenvalues $|A| = \prod_{i=1}^n \lambda_i$.
- The rank of A is equal to the number of non-zero eigenvalues of A.
- If A is non-singular then $1/\lambda_i$ is an eigenvalue of A^{-1} with associated eigenvector x_i i.e., $A^{-1}x_i = (1/\lambda_i)\, x_i$. (To prove this, take the eigenvector equation, $Ax_i = \lambda_i x_i$ and left-multiply each side by A^{-1}.)
- The eigenvalues of a diagonal matrix $D = \operatorname{diag}(d_1, \ldots d_n)$ are just the diagonal entries $d_1, \ldots d_n$.

We can write all the eigenvector equations simultaneously as $AX = X\Lambda$, where

$$X \in \mathbb{R}^{n\times n} = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix},\ \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \tag{24}$$

If the eigenvectors of A are i.i.d, then X will be invertible, so $A = X\Lambda X^{-1}$. A matrix that can be written in this form is called **diagonalizable**.

Eigenvalues & Eigenvectors of Symmetric Matrics For a symmetric matrix $A \in \mathcal{S}^n$, we have

- All the eigenvalues of A are real.
- The eigenvectos of A are orthnormal, i.e., X is an orthogonal matrix, also denoted as U. Thus we have $A = X\Lambda X^{-1} = U\Lambda U^T$.

Using this, we can show that the definiteness of a matrix depends entirely on the sign of its eigenvalues, for

$$x^T Ax = x^T U\Lambda U^T x = y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2. \tag{25}$$

The definiteness of matrix A depends only on λ_i.

Maximizing of Matrices Given $A \in \mathcal{S}^n$, consider

$$\max_{x\in\mathbb{R}^n} x^T Ax \quad \text{subject to } \|x\|_2^2 = 1. \tag{26}$$

If λ_i are ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, the optimal x for Eqn 26 is x_1, the eigenvector corresponding to λ_1. The maximal value of the quadratic form is λ_1.

1.2 Matrix Calculus

Gradient of Matrices Suppose that $f : \mathbb{R}^{m\times n} \to \mathbb{R}$,

$$\nabla_A f(A) \in \mathbb{R}^{m\times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix} \tag{27}$$

where

$$(\nabla_A f(A))_{ij} = \frac{\partial f(A)}{\partial A_{ij}}. \tag{28}$$

Properties

- $\nabla_x(f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x)$
- For $t \in \mathbb{R}, \nabla_x(tf(x)) = t\nabla_x f(x)$

Hessian Matrix Suppose that $f : \mathbb{R}^{m\times n} \to \mathbb{R}$,

$$\nabla_x^2 f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \tag{29}$$

where

$$\left(\nabla_x^2 f(x)\right)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}. \tag{30}$$

The Hessian is always symmetric.

If $f(x) = x^T Ax$ for $A \in \mathcal{S}^n$, remember that $f(x) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$, we have

$$\begin{aligned}
\frac{\partial f(x)}{\partial x_k} &= \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j \\
&= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right] \\
&= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k \\
&= \sum_{i=1}^n A_{ik} x_i + \sum_{j=1}^n A_{kj} x_j = 2 \sum_{i=1}^n A_{ki} x_i
\end{aligned} \tag{31}$$

Thus we have $\nabla_x x^T Ax = 2Ax$.

For Hessian,

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_\ell} = \frac{\partial}{\partial x_k} \left[\frac{\partial f(x)}{\partial x_\ell} \right] = \frac{\partial}{\partial x_k} \left[2 \sum_{i=1}^n A_{\ell i} x_i \right] = 2A_{\ell k} = 2A_{k\ell}. \tag{32}$$

where $\nabla_x^2 x^T Ax = 2A$.

To recap,

- $\nabla_x b^T x = b$
- $\nabla_x x^T Ax = 2Ax$ (if A symmetric)
- $\nabla_x^2 x^T Ax = 2A$ (if A symmetric)

Least Squares For $A \in \mathbb{R}^{m \times n}$ (for simplicity we assume A is full rank) and $b \in \mathbb{R}^m$ such that $b \notin \mathcal{R}(A)$. We have

$$\begin{aligned}
\|Ax - b\|_2^2 &= (Ax - b)^T (Ax - b) \\
&= x^T A^T Ax - 2b^T Ax + b^T b
\end{aligned} \tag{33}$$

The gradient will be

$$\begin{aligned}
\nabla_x \left(x^T A^T Ax - 2b^T Ax + b^T b \right) &= \nabla_x x^T A^T Ax - \nabla_x 2b^T Ax + \nabla_x b^T b \\
&= 2A^T Ax - 2A^T b
\end{aligned} \tag{34}$$

Thus

$$x = \left(A^T A \right)^{-1} A^T b. \tag{35}$$

Need to clarify $\nabla_x 2b^T Ax = 2A^T b$.

Gradients of the Determinant For $A \in \mathbb{R}^{n \times n}$, as $|A| = \sum_{i=1}^n (-1)^{i+j} a_{ij} \left| A_{-i,-j} \right|$, we have

$$\frac{\partial}{\partial A_{k\ell}} |A| = \frac{\partial}{\partial A_{k\ell}} \sum_{j=1}^n (-1)^{i+j} A_{ij} \left| A_{-i,-j} \right| = (-1)^{k+\ell} \left| A_{-k,-\ell} \right| = (\operatorname{adj}(A))_{\ell k}. \tag{36}$$

Thus

$$\nabla_A |A| = (\operatorname{adj}(A))^T = |A| A^{-T}. \tag{37}$$

Eigenvalues as Optimization Recall the constrained optimization problem:

$$\max_{x \in \mathbb{R}^n} x^T A x \quad \text{subject to } \|x\|_2^2 = 1 \tag{38}$$

for a symmetric matrix $A \in \mathbb{S}^n$. We can use **Lagrangian** to optimize:

$$\mathcal{L}(x, \lambda) = x^T A x - \lambda x^T x \tag{39}$$

where λ is called the Lagrange multiplier associated with the equality constraint. The gradient of the Lagrangian has to be zero at $x^\star$, that is

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla_x \left(x^T A x - \lambda x^T x \right) = 2A^T x - 2\lambda x = 0. \tag{40}$$

Notice that this is just the linear equation $Ax = \lambda x$. This shows that the only points which can possibly maximize (or minimize) $x^T A x$ assuming $x^T x = 1$ are the eigenvectors of A.

References

[1] Zico Kolter. Linear algebra review and reference. 2008.