Machine Learning in Computer Vision: SPD Matrices and Norms

Oren Freifeld
Computer Science, Ben-Gurion University

March 8, 2018
**Definition (positive-definite matrix)**

An $n \times n$ matrix, $A$, is called Positive Definite (PD) if

$$x^T A x > 0 \quad \forall \text{ non-zero } x \in \mathbb{R}^n .$$  \hfill (1)

**Example**

$I_{n \times n}$, the identity matrix, is SPD:

$$x^T I x = x^T x = \sum_{i=1}^{n} x_i^2 > 0 \quad \forall \text{ non-zero } x \in \mathbb{R}^n .$$  \hfill (2)

**Definition (positive-semidefinite matrix)**

An $n \times n$ matrix, $A$, is called Positive Semidefinite (PSD) if

$$x^T A x \geq 0 \quad \forall \text{ non-zero } x \in \mathbb{R}^n .$$  \hfill (3)
**Definition (symmetric matrix)**

An $n \times n$, $A$, is called symmetric if $A = A^T$.

**Definition (symmetric positive-definite matrix)**

An $n \times n$, $A$ is called SPD if it is both symmetric and PD.

**Definition (symmetric positive-semidefinite matrix)**

An $n \times n$, $A$ is called SPSD if it is both symmetric and PSD.

**Fact**

- A symmetric matrix $A$ is SPD $\iff$ all its eigenvalues are positive $\iff$ it has a unique Cholesky decomposition; namely, there exists a unique lower triangular matrix $L$, with positive diagonal elements, such that $A = LL^T$.
- A symmetric matrix $A$ is PSD $\iff$ all its eigenvalues are nonnegative.
Definition ($\ell_p$ norm (for $n$-tuples))

For $\mathbf{x} = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$, 

$$\|\mathbf{x}\|_{\ell_p} \triangleq \left( \sum_{j=1}^{n} |x_j|^p \right)^{1/p} \quad 1 \leq p < \infty$$

(4)

is the $\ell_p$ norm of $\mathbf{x}$.

Example ($\ell_2$ norm – AKA the Euclidean norm (for $n$-tuples))

$$\|\mathbf{x}\|_{\ell_2}^2 \triangleq \sum_{j=1}^{n} x_j^2$$

(5)

is the squared $\ell_2$ norm of $\mathbf{x}$. 
Example ($\ell_1$ norm (for $n$-tuples))

$$
\|x\|_{\ell_1} \triangleq \sum_{j=1}^{n} |x_j|
$$

(6)

is the $\ell_1$ norm of $x$. 
Definition ($\ell_p$ distance (for $n$-tuples))

For $x, y \in \mathbb{R}^n$,

$$\|x - y\|_{\ell_p} \quad 1 \leq p < \infty$$

(7)

is the $\ell_p$ distance of between $x$ and $y$.

Example ($\ell_2$ distance – AKA the Euclidean distance (for $n$-tuples))

$$d^2(x, y) = \|x - y\|_{\ell_2}^2$$

(8)

is the squared $\ell_2$ distance between $x$ and $y$.

Since Euclidean distance is usually the default, often one drops the subscript and just writes $\|x - y\|$ instead of $\|x - y\|_{\ell_2}$. 
Example ($\ell_1$ distance (for $n$-tuples))

\[ d(x, y) = \| x \|_{\ell_1} \triangleq \sum_{j=1}^{n} |x_j - y_j| \]  

(9)

is the $\ell_1$ distance between $x$ and $y$.

Remark

The norms and distances above are easily modified for the case of infinite sequences instead of $n$-tuples. For example, the $\ell_2$ norm for sequences of the form $(x_1, x_2, \ldots)$ is given by

\[ \| x \|_{\ell_2}^2 \triangleq \sum_{j=1}^{\infty} x_j^2. \]
Example (norm and distance defined via an SPD matrix)

Let $Q \in \mathbb{R}^{n \times n}$ be SPD. Then we let

$$\|x\|_Q^2 \triangleq x^T Q x$$

and

$$d_Q^2(x, y) = \|x - y\|_Q^2 = (x - y)^T Q (x - y)$$

be the norm and induced distance defined by $Q$.

Let $Q$ be an SPD matrix and $LL^T$ denote its (unique) Cholesky decomposition. Note that

$$\|x\|_Q^2 = x^T Q x = x^T L L^T x = (L^T x)^T L^T x = \|L^T x\|_{\ell_2}^2.$$