Lecture 4 : Metrics from matrices

updated 30/7/21

We have still not properly analysed the example of $\mathbb{R}^n$ with the Euclidean distance d_2 as a metric space, i.e.

$$d_2(\underline{x},\underline{y}) = \left\{ \sum_{i=1}^n (x_i - y_i)^2 \right\}^{1/2}$$
$$= \left\{ (\underline{x}-\underline{y})^T(\underline{x}-\underline{y}) \right\}^{1/2}$$

In this lecture we return to this example, in the context of a more general discussion of deriving metrics d_B on $\mathbb{R}^n$ from special kinds of matrices $B \in M_n(\mathbb{R})$. Here $M_n(\mathbb{R})$ denotes the set of all $n \times n$ matrices. But first, let us (finally) actually check this is a metric space:

Lemma L4-1 $(\mathbb{R}^n, d_2)$ is a metric space

Proof (M1)-(M3) are clear. For (M4) we first prove

Claim For $r_1,\dots,r_n, s_1,\dots,s_n \in \mathbb{R}$ $\quad \sum_i r_i^2 \sum_i s_i^2 \geq \left(\sum_i r_i s_i\right)^2$. (Cauchy's inequality)

Proof Set $f: \mathbb{R} \to \mathbb{R}$ by $f(\lambda) = \sum_i (r_i + \lambda s_i)^2$. Then $f(\lambda) \geq 0$ for all λ and f is quadratic, so we know its discriminant must be ≤ 0. But the discriminant of

$$f(\lambda) = \sum_i \left(s_i^2 \lambda^2 + 2\lambda r_i s_i + r_i^2 \right)$$

is $\Delta = 4\left(\sum_i r_i s_i\right)^2 - 4\sum_i s_i^2 \sum_j r_j^2$ and so $\Delta \leq 0$ is what we wanted to show. $\square$

To prove (M4) we have to show

$$d_2(\underline{x},\underline{y}) + d_2(\underline{y},\underline{z}) \geqslant d_2(\underline{x},\underline{z})$$

which means

$$\left\{ \sum_i (x_i - y_i)^2 \right\}^{1/2} + \left\{ \sum_i (y_i - z_i)^2 \right\}^{1/2} \geqslant \left\{ \sum_i (x_i - z_i)^2 \right\}^{1/2}.$$

Set $a_i = x_i - y_i$ and $b_i = y_i - z_i$ so that $a_i + b_i = x_i - z_i$. Then

$$\begin{aligned}
\sum_i (a_i + b_i)^2 &= \sum_i (a_i^2 + 2a_i b_i + b_i^2) \\
&= \sum_i a_i^2 + 2\sum_i a_i b_i + \sum_i b_i^2 \\
&\overset{\text{Cauchy}}{\leq} \sum_i a_i^2 + 2\left(\sum_i a_i^2\right)^{1/2}\left(\sum_i b_i^2\right)^{1/2} + \sum_i b_i^2 \\
&= \left\{ \left(\sum_i a_i^2\right)^{1/2} + \left(\sum_i b_i^2\right)^{1/2} \right\}^2
\end{aligned}$$

taking the square root on both sides completes the proof of (M4). □

<u>Exercise L4-0</u> Prove that

$$d_1(\underline{x},\underline{y}) = \sum_{i=1}^{n} |x_i - y_i|$$

$$d_\infty(\underline{x},\underline{y}) = \max\{ |x_i - y_i| \}_{i=1}^{n}$$

define metrics on $\mathbb{R}^n$.

<u>Exercise L4-1</u> If (X, d_X) is a metric space and $\rho : Y \longrightarrow X$ is a bijection (Y any set) then (Y, d_ρ) is a metric space where

$$d_\rho(y_1, y_2) := d_X(\rho(y_1), \rho(y_2)).$$

<u>Lemma L4-2</u> Let $B \in M_n(\mathbb{R})$ be an invertible matrix. Then

$$d_B(\underline{v}, \underline{w}) = \{ (\underline{v}-\underline{w})^T B^T B (\underline{v}-\underline{w}) \}^{1/2}$$

is a metric on $\mathbb{R}^n$.

<u>Proof</u> Let $\rho : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be left multiplication by B. Then d_ρ as above is a metric, and

$$\begin{aligned} d_\rho(\underline{v}, \underline{w}) &= d_2(B\underline{v}, B\underline{w}) \\ &= \{ (B\underline{v} - B\underline{w})^T (B\underline{v} - B\underline{w}) \}^{1/2} \\ &= \{ [B(\underline{v}-\underline{w})]^T [B(\underline{v}-\underline{w})] \}^{1/2} \\ &= \{ (\underline{v}-\underline{w})^T B^T B (\underline{v}-\underline{w}) \}^{1/2}. \quad \square \end{aligned}$$

<u>Example L4-0</u> If $\lambda_1, \ldots, \lambda_n > 0$ then

$$d(\underline{x}, \underline{y}) = \{ \textstyle\sum_{i=1}^n \lambda_i (x_i - y_i)^2 \}^{1/2}$$

is a metric on $\mathbb{R}^n$ (take $B = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$). This metric corresponds to "shrinking" our measuring stick for axis i (if $\lambda_i > 1$) or "stretching" it (if $\lambda_i < 1$).

<u>Question</u> For which matrices P is $(\underline{v},\underline{w}) \mapsto \{(\underline{v}-\underline{w})^T P(\underline{v}-\underline{w})\}^{1/2}$ a metric?

NOTE we know $P = B^T B$ for B invertible works

Let us see what constraints the metric axioms impose on P

(M1) will hold iff. $\underline{x}^T P \underline{x} \geqslant 0 \quad \forall \underline{x} \in \mathbb{R}^n$

(M2) will hold iff. $\underline{x}^T P \underline{x} = 0 \iff \underline{x} = 0$.

(M3) will hold automatically since

$$(\underline{v}-\underline{w})^T P(\underline{v}-\underline{w}) = (\underline{w}-\underline{v})^T P(\underline{w}-\underline{v}).$$

(M4) Recall that for d_2, which is the case $P = I_n$, we deduced the triangle inequality from Cauchy's inequality

$$\textstyle\sum_i r_i^2 \sum_i s_i^2 \geqslant \left(\sum_i r_i s_i\right)^2.$$

We want to now generalise this. Define for $\underline{v},\underline{w} \in \mathbb{R}^n$

$$\langle \underline{v},\underline{w}\rangle_P := \underline{v}^T P \underline{w} \in \mathbb{R}$$

Note that if $P = P^T$ is symmetric then $\langle \underline{v},\underline{w}\rangle_P = \langle \underline{w},\underline{v}\rangle_P$, is also symmetric. Notice that with $P^{sym} := \frac{1}{2}(P + P^T)$ we have

$$\begin{aligned} \underline{x}^T P^{sym} \underline{x} &= \tfrac{1}{2}\underline{x}^T P \underline{x} + \tfrac{1}{2} \underline{x}^T P^T \underline{x} \\ &= \tfrac{1}{2}\underline{x}^T P \underline{x} + \tfrac{1}{2}(\underline{x}^T P \underline{x})^T = \underline{x}^T P \underline{x} \end{aligned}$$

so if we use P or P^{sym} we obtain the same (candidate) metric. So we may without loss of generality assume P is symmetric.

OK, so the only <u>potential</u> metrics obtained in this way from matrices P are obtained from the following type of matrix:

<u>Def^n</u> A matrix $P \in M_n(\mathbb{R})$ is called <u>positive-definite</u> if it is

(i) symmetric, $P = P^T$

(ii) $\underline{x}^T P \underline{x} \geqslant 0$ for all $\underline{x} \in \mathbb{R}^n$ (positive)

(iii) $\underline{x}^T P \underline{x} = 0 \iff \underline{x} = 0$ (definite)

<u>Exercise L4-2</u> Prove that $\langle -,- \rangle_P$ is linear in both variables (we say <u>bilinear</u>) by which we mean

$$\langle \lambda_1 \underline{v}_1 + \lambda_2 \underline{v}_2 , \underline{w} \rangle_P = \lambda_1 \langle \underline{v}_1, \underline{w} \rangle_P + \lambda_2 \langle \underline{v}_2, \underline{w} \rangle_P,$$
$$\langle \underline{v}, \lambda_1 \underline{w}_1 + \lambda_2 \underline{w}_2 \rangle_P = \lambda_1 \langle \underline{v}, \underline{w}_1 \rangle_P + \lambda_2 \langle \underline{v}, \underline{w}_2 \rangle_P$$

<u>Lemma L4-3</u> (Cauchy-Schwartz, partial version) For any positive-definite $P \in M_n(\mathbb{R})$ and any $\underline{v}, \underline{w} \in \mathbb{R}^n$,

$$\langle \underline{v}, \underline{w} \rangle_P^2 \leq \langle \underline{v}, \underline{v} \rangle_P \cdot \langle \underline{w}, \underline{w} \rangle_P$$

<u>Proof</u> The trick is similar to what we did before. Observe that for any $\lambda \in \mathbb{R}$, writing $\langle , \rangle$ for $\langle , \rangle_P$

$$\begin{aligned} 0 &\leq \langle \underline{v} - \lambda \underline{w}, \underline{v} - \lambda \underline{w} \rangle \\ &= \langle \underline{v}, \underline{v} \rangle - \lambda \langle \underline{v}, \underline{w} \rangle - \lambda \langle \underline{w}, \underline{v} \rangle + \lambda^2 \langle \underline{w}, \underline{w} \rangle \\ \text{using } P \text{ symmetric } &= \langle \underline{v}, \underline{v} \rangle - 2\lambda \langle \underline{v}, \underline{w} \rangle + \lambda^2 \langle \underline{w}, \underline{w} \rangle. \end{aligned}$$

The desired identity is trivial if $\underline{w}=0$ so assume not, and set

$$\lambda = \langle \underline{v}, \underline{w} \rangle / \langle \underline{w}, \underline{w} \rangle$$

Here we use that P is definite so $\langle \underline{w}, \underline{w} \rangle \neq 0$. Then

$$0 \leq \langle \underline{v}, \underline{v} \rangle - \frac{2 \cdot \langle \underline{v}, \underline{w} \rangle^2}{\langle \underline{w}, \underline{w} \rangle} + \frac{\langle \underline{v}, \underline{w} \rangle^2}{\langle \underline{w}, \underline{w} \rangle}$$

$$= \langle \underline{v}, \underline{v} \rangle - \frac{\langle \underline{v}, \underline{w} \rangle^2}{\langle \underline{w}, \underline{w} \rangle}$$

Multiplying through by $\langle \underline{w}, \underline{w} \rangle$, and using that P is positive (so that this does not reverse the inequality) we have

$$\langle \underline{v}, \underline{w} \rangle^2 \leq \langle \underline{v}, \underline{v} \rangle \cdot \langle \underline{w}, \underline{w} \rangle. \quad \square$$

<u>Proposition L4-4</u> If $P \in M_n(\mathbb{R})$ is positive definite then

$$d_P(\underline{x}, \underline{y}) = \{ (\underline{x}-\underline{y})^T P (\underline{x}-\underline{y}) \}^{1/2}$$

is a metric on $\mathbb{R}^n$.

<u>Proof</u> We need only prove (M4). But again given $\underline{x}, \underline{y}, \underline{z}$ set $\underline{a} = \underline{x} - \underline{y}$ and $\underline{b} = \underline{y} - \underline{z}$ so we need to show

$$\langle \underline{a}, \underline{a} \rangle_P^{1/2} + \langle \underline{b}, \underline{b} \rangle_P^{1/2} \geq \langle \underline{a}+\underline{b}, \underline{a}+\underline{b} \rangle_P^{1/2}$$

Calculating using the previous lemma:

$$\left\{ \langle \underline{a},\underline{a}\rangle_P^{1/2} + \langle \underline{b},\underline{b}\rangle_P^{1/2} \right\}^2 = \langle \underline{a},\underline{a}\rangle_P + 2\langle \underline{a},\underline{a}\rangle_P^{1/2}\langle \underline{b},\underline{b}\rangle_P^{1/2} + \langle \underline{b},\underline{b}\rangle_P$$

$$= \langle \underline{a},\underline{a}\rangle_P + 2\left\{ \langle \underline{a},\underline{a}\rangle_P\langle \underline{b},\underline{b}\rangle_P \right\}^{1/2} + \langle \underline{b},\underline{b}\rangle_P$$

$$\geq \langle \underline{a},\underline{a}\rangle_P + 2\left\{ \langle \underline{a},\underline{b}\rangle_P^2 \right\}^{1/2} + \langle \underline{b},\underline{b}\rangle_P$$

$$\geq \langle \underline{a},\underline{a}\rangle_P + 2\langle \underline{a},\underline{b}\rangle_P + \langle \underline{b},\underline{b}\rangle_P$$

$$= \langle \underline{a}+\underline{b},\underline{a}+\underline{b}\rangle_P .$$

Taking square-roots proves the claim. $\square$

<u>Exercise L4-3</u> Prove that if $B \in M_n(\mathbb{R})$ is invertible then $B^T B$ is positive-definite.

<u>Exercise L4-4</u> Prove that if $P_1 = Q^{-1} P_2 Q$ for some orthogonal matrix Q then multiplication by Q gives an isometry (assume P_1, P_2 positive-definite)

$$(\mathbb{R}^n, d_{P_1}) \longrightarrow (\mathbb{R}^n, d_{P_2}).$$

That is, the metric we get on $\mathbb{R}^n$ from P_1 is <u>essentially the same</u> as the one we get from P_2.

<u>Exercise L4-5</u> Prove that every real symmetric matrix X can be written as $X = Q^T D Q$ with Q orthogonal and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ diagonal, and that X is positive definite iff. all the λ_i are positive.

<u>Conclusion</u> essentially the only new examples of metrics on $\mathbb{R}^n$ that we obtain in this way are those obtained from matrices

$$P = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \qquad \text{all } \lambda_i > 0.$$

In which case

$$d_P(\underline{x}, \underline{y}) = \left\{ (\underline{x}-\underline{y})^T P (\underline{x}-\underline{y}) \right\}^{1/2}$$

$$= \left\{ \sum_{i=1}^n \lambda_i (x_i - y_i)^2 \right\}^{1/2}$$

But this was the example we say already in Example L4-0! So we failed! It was completely pointless to explore this idea of inducing new metrics from matrices! We only found things isometric to an example we already knew!

Well, that's all true, but in the process of failing we acquired something very valuable: a new point of view on metrics as arising from matrices via $\underline{v}^T P \underline{w}$, which puts us on the cusp of discovering a remarkable class of spaces which represent a radical departure from our ordinary (= limited, parochial, narrow) intuitions about space based on $\mathbb{R}^n$. To make this discovery we need only ask the following

<u>Question</u> What happens if some of the λ_i in $P = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ are <u>negative</u>? What is the geometric content of $\langle -, - \rangle_P$?