*Acta Cryst.* (1997). A**53**, 809

# Metric properties of the root-mean-square deviation of vector sets

Karl Kaindl[a]* and Boris Steipe[b] *at* [a]*Department of Mathematics and Genzentrum, 81377 Munich, Feodor-Lynen-Strasse, Ludwig-Maximilians-University, Munich, Germany, and* [b]*Genzentrum, Ludwig-Maximilians-University, Munich, Germany. E-mail: kaindl@rz.mathematik.uni-muenchen.de*

*(Received 14 June 1997; accepted 18 July 1997)*

**Abstract**

A proof is given that the root-mean-square deviation between more than two vector sets after optimal superposition induces a metric.

---

The RMSD (root-mean-square deviation) is commonly used as a criterion to measure the similarity of two $n$-dimensional vector sets (*e.g.* molecular structures) after optimal superposition. An analytic solution to the calculation of the best rotation matrix that minimizes the RMSD is given by Kabsch (1976). If more than two structure sets are to be compared, the question arises whether the RMSD's fulfil the requirements to be a metric in the mathematical sense. The metric properties of measures on elements of data sets are often implied in data-analysis procedures like cluster analysis or sampling of representatives. For a measure to be a metric $d$ on a set $X$, $d : X \times X \to R_+$ we require

(i) $\forall x, y \in X: \quad d(x, y) \geq 0, \quad d(x, y) = 0 \Leftrightarrow x = y$ (positivity),

(ii) $d(y, x) = d(x, y)$ (symmetry),

(iii) $\forall z \in X: \quad d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

If conditions (i)–(iii) hold, we call $d$ a metric on a space $X$ and call the pair $(X, d)$ a metric space. We specify $X := R^{3,n}$, $x, y, z \in R^{3,n}$ to be vector sets. Since the calculation of the RMSD is performed on vector sets with the same number of $n$ elements, we may ignore the common dividing factor by multiplying the RMSD with $n$. Thus, the RMSD is equivalent to the Euclidean norm in $R^{3,n}$. Let $d$ be a mapping

$$d : R^{3,n} \times R^{3,n} \to R_+, \quad d(x, y) := ||x - y||.$$

We can rewrite (i)–(iii) as

(i′) $||x - y|| \geq 0, \quad ||x - y|| = 0 \Leftrightarrow x = y$,

(ii′) $||x - y|| = ||y - x||$,

(iii′) $||x - z|| \leq ||x - y|| + ||y - z||$.

Obviously, the RMSD of vector sets is a metric since $d$ defines a metric on the Euclidean space. But it is not obvious whether $\mathrm{RMSD}_{\mathrm{opt}}$, the RMSD after optimal superposition, is a metric. Since all vector sets that differ only by an arbitrary rotation and translation can be seen as elements of an equivalence class, the $\mathrm{RMSD}_{\mathrm{opt}}$ is the solution of a minimization problem for all pairs of representatives of such classes, even though for each two classes a unique minimum exists (Stoer, 1983). Thus, $\mathrm{RMSD}_{\mathrm{opt}}$ defines a relationship between representatives of equivalence classes of vector sets.

After optimal superposition, the vector set centroids coincide. Thus, a translation can always be removed by shifting the vector centroids to the origin of the coordinate system (Kabsch, 1976). Let $M(x, y) \in R^{3,3}$ be a rotation matrix.† We set $\mathrm{RMSD}_{\mathrm{opt}}(x, y) := \min ||y - M(x, y)x||$ and $\tilde{x}_y := M(x, y)x \Leftrightarrow \min ||y - M(x, y)x||$ (*etc.* for $y$ and $z$). Since the rotations are different for each pair of vector sets, generally $M(x, z) \neq M(y, z) \circ M(x, y)$, which would make things much easier.

If two vector sets differ only by a rotation and translation, we may define the following relation on $R^{3,n}$:

$$x \sim y :\Leftrightarrow ||x - \tilde{y}_x|| = 0$$

$\forall x, y \in R^{3,n}$. This relation defines an equivalence relation and we write

$$[x] := \{y \in R^{3,n} : x \sim y\}$$

for the equivalence class $[x]$. For any $x \in [x]$, the $\mathrm{RMSD}_{\mathrm{opt}}(x, y)$ is equal to the RMSD $(\tilde{x}_y, y)$, thus $\tilde{x}_y$ is taken as representative of $[x]$. Condition (i′) states that $||x - \tilde{y}_x|| \geq 0$; $||x - \tilde{y}_x|| = 0 \Leftrightarrow x \sim \tilde{y}_x \Leftrightarrow y \in [x]$. For reasons of symmetry, we have in (ii′) $||x - \tilde{y}_x|| = ||y - \tilde{x}_y||$.

Now we have to show (iii′), which is more difficult.

Since $||z - \tilde{x}_z|| \leq ||z - x||$ and $||y - \tilde{x}_y|| \leq ||y - x||$ hold and the RMSD is a metric on $R^{3,n}$, we can apply the triangle inequality and get

$$\begin{aligned} ||z - \tilde{x}_z|| \leq ||z - x|| &= ||z - \tilde{y}_z + \tilde{y}_z - x|| \\ &\leq ||z - \tilde{y}_z|| + ||\tilde{y}_z - x|| \end{aligned} \quad (1)$$

$$\begin{aligned} ||y - \tilde{x}_y|| \leq ||y - x|| &= ||y - \tilde{y}_z + \tilde{y}_z - x|| \\ &\leq ||y - \tilde{y}_z|| + ||\tilde{y}_z - x||. \end{aligned} \quad (2)$$

If we subtract (2) from (1), we get

$$||z - \tilde{x}_z|| \leq ||y - \tilde{x}_y|| + ||z - \tilde{y}_z|| - ||y - \tilde{y}_z||.$$

From (ii′) and since $-||y - \tilde{y}_z|| \leq 0$,

$$||x - \tilde{z}_x|| \leq ||x - \tilde{y}_x|| + ||y - \tilde{z}_y||$$

follows, which is (iii′).

† For all rotation matrices $||M \circ x|| = ||x||$ holds.

**References**

Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.

Stoer, J. (1983). *Numerische Mathematik I*, p. 181. Heidelberg: Springer Verlag.

Acta Crystallographica Section A
**Foundations of Crystallography**
ISSN 0108-7673

# addenda and errata

## Metric properties of the root-mean-square deviation of vector sets. Erratum

**Boris Steipe**

Department of Biochemistry and Program in Proteomics and Bioinformatics, University of Toronto, Toronto, Ontario, Canada

A technical error invalidates the proof that the optimal root-mean-square deviation of vector sets induces a metric given in Kaindl & Steipe [*Acta Cryst*. (1997), A**53**, 809]. Nevertheless, the conclusions are correct and a revised proof is given in Steipe [*Acta Cryst.* (2002), A**58**, 506].

The author is indebted to Dr Xu Huafeng for pointing out the technical error in the previously published manuscript (Kaindl & Steipe, 1997).

### References

Kaindl, K. & Steipe, B. (1997). *Acta Cryst.* A**53**, 809.
Steipe, B. (2002). *Acta Cryst.* A**58**, 506.

## A multigrid approach to the average lattices of quasicrystals. Erratum

**J. L. Aragón,[a]* Gerardo G. Naumis[b] and M. Torres[c]**

[a]Instituto de Física, Universidad Nacional Autónoma de México, Apartado Postal 1-1010, 76000 Querétaro, Mexico, [b]Instituto de Física, Universidad Nacional Autónoma de México, Apartado Postal 20-364, 01000 Distrito Federal, Mexico, and [c]Instituto de Física Aplicada, Consejo Superior de Investigaciones Científicas, Serrano 144, 28006 Madrid, Spain

The corresponding author's e-mail address was wrongly given in the paper by Aragón *et al*. [*Acta Cryst.* (2002), A**58**, 352–360]. The correct address is aragon@fata.unam.mx.

### References

Aragón, J. L., Naumis, G. G. & Torres, M. (2002). A**58**, 352–360.

  