DNA Modelling Course
Exercise Session 1
Summer 2006

SOLUTIONS

1 The length scales of DNA

The genome of an organism is the sequence of DNA base-pairs (A, T, C and G) that this organism contains. The DNA in human cells is organised in macromolecules called chromosomes. The human genome contains 22 chromosomes plus 2 sexual chromosomes, called X and Y. Its length is nowdays estimated at $3,2 \cdot 10^9$ base-pairs.
You can see the genome map on the website of the Human Genome project at http://www.ornl.gov/TechResources/Human_Genome/posters/chromosome/

The large majority of human cells are diploids. It means that they contain two copies of each non-sexual chromosome plus two sexual chromosomes, that is, $2 \cdot 22 + 2 = 46$ chromosomes. Thus, a diploid cell contains two times the genome, that is, approximately $6,4 \cdot 10^9$ base pairs of DNA.

(a) If T is the total number of base-pairs in DNA and h is the base-pair length, then the total length L of DNA in a cell is given by

$$\begin{aligned} L &= T \cdot h \\ &= 6 \cdot 10^9 \cdot 3,4 \cdot 10^{-10}\,\mathrm{m} = 6 \cdot 3,4 \cdot 10^{-1}\,\mathrm{m} \\ &= 2.04\,\mathrm{m} \qquad (1.1) \end{aligned}$$

Thus the total length of DNA in a cell is approximately 2 meters, or $\cdot 10^5$ times the diameter of the cell. The DNA must therfore be highly folded!

(b)

$$r = \frac{(\text{DNA diameter})}{2} = \frac{20}{2}\text{Å} = 10^{-9}\text{m} \qquad (1.2)$$

$$R = \frac{(\text{cell diameter})}{2} = \frac{10^5}{2}\text{Å} = 5 \cdot 10^{-6}\text{m} \qquad (1.3)$$

The volume of DNA is given by

$$\begin{aligned} v &= L \cdot \pi \cdot r^2 \\ &= 2,04 \cdot \pi \cdot 10^{-18}\,\mathrm{m}^3 \\ &= 6,41 \cdot 10^{-18}\mathrm{m}^3 \end{aligned} \tag{1.4}$$

The volume of the cell is given by

$$\begin{aligned} V &= \frac{4}{3} \cdot \pi \cdot R^3 \\ &= \frac{4}{3} \cdot \pi \cdot 1,25 \cdot 10^{-16}\,\mathrm{m}^3 \\ &= 5,23 \cdot 10^{-16}\,\mathrm{m}^3 \end{aligned} \tag{1.5}$$

so that

$$\frac{V}{v} = 81,68 \tag{1.6}$$

The DNA volume is thus 81 times smaller than the cell volume! That is it fills only about 1.22% of the available volume. (In point of fact in eukaryotes the DNA is stored inside the nucleus of the cell so that the available volume is substantially smaller than that of the cell itself. But nevertheless there is plenty of volume available to store the DNA—the problem is how to coil such a long object in an organized fashion to be packed.)

For those who are interested in more details about the human genome: The human genome sequence as been published in Science [1] in 2001. 2.91-billion base pair (bp) of the euchromatic portion of the human genome were sequenced, including the sequence of 22 chromosomes and the sexual chromosomes X (155 millions bp) and Y (58 millions bp) [2].
A typical human cell is dioploidic which means it has two copies of each non-sexual chromosome, so it has for a male $(2.91 - \mathrm{x} - \mathrm{y}) \cdot 2 + \mathrm{x} + \mathrm{y} = 2 \cdot 2.91 - \mathrm{x} - \mathrm{y} = (2 \cdot 2.9$ billions-155 millions -58millions$=$ 5587 millions base pairs of DNA, and for a female $(2.91 - \mathrm{x} - \mathrm{y}) \cdot 2 + 2\mathrm{x} = 2 \cdot 2.91 - 2\mathrm{y} = 5684$ millions bp.
The cell also contains 10% of another type of DNA called heterochromatin[3]

[1] The Sequence of the Human Genome Science 16 February 2001 Vol. 291. no. 5507, pp. 1304 - 1351

[2] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=9558

[3] Alberts et al, Molecular Biology of the Cell third edition, Garland Publishing, 1994, pp 353

that was not sequenced in the Science publication, so the total number of base pairs has to be increased by ten percent: $5.6 \cdot 1.1$ billions ≈ 6 billions bp.
Before mitosis (cell division) all the DNA is replicated, which doubles the total amount of DNA in the cell (around 12 billion bp).
A base pair junction is 3.4 Angstrom ($0.34 \cdot 10 - 9$m) long so the total DNA length varies between $6 \cdot 10^9 \cdot 0.34 \cdot 10^{-9}$ and $12 \cdot 10^9 \cdot 0.34 \cdot 10^{-9}$ meters, that is between 2 to 4 meters depending on the phase in the cell cycle.
The cell has also mitochondrial DNA in the cytoplasm but it is negligible (only a few tens of thousands bp).

2 Rotations in three dimensions

(a) From the properties of the scalar product, $\forall \boldsymbol{w} \in I\!R^3$

$$\begin{aligned} \|\boldsymbol{Qw}\|^2 &= \boldsymbol{Qw} \cdot \boldsymbol{Qw} \\ &= \boldsymbol{w} \cdot \boldsymbol{Q}^T \boldsymbol{Qw} \\ &= \|\boldsymbol{w}\|^2 \end{aligned} \tag{2.1}$$

where we have used the fact that $\boldsymbol{Q}$ is a rotation matrix, i.e. $\boldsymbol{Q}^T\boldsymbol{Q} = I$. If now λ is an eigenvalue for $\boldsymbol{Q}$, by definition there must exist a vector $\boldsymbol{w}$, which is the corresponding eigenvector, so that the following equivalences hold

$$\begin{aligned} \|\boldsymbol{Qw}\| &= \|\lambda \boldsymbol{w}\| \\ &= |\lambda| \|\boldsymbol{w}\| \end{aligned} \tag{2.2}$$

but then from equation 2.1 we obtain

$$|\lambda| = 1 \tag{2.3}$$

and therefore λ lies on the unit circle in the complex plane.

(b) Note that the complex conjugate $\overline{\lambda}$ is an eigenvalue of $\boldsymbol{Q}$ (with corresponding eigenvector $\overline{\boldsymbol{w}}$) whenever λ is an eigenvalue of $\boldsymbol{Q}$ (with corresponding eigenvector $\boldsymbol{w}$). This follows from $\overline{\boldsymbol{Q}} = \boldsymbol{Q}$, explicitly,

$$\boldsymbol{Qw} = \lambda \boldsymbol{w} \;\Rightarrow\; \boldsymbol{Q}\overline{\boldsymbol{w}} = \overline{\boldsymbol{Qw}} = \overline{\lambda \boldsymbol{w}} = \overline{\lambda}\,\overline{\boldsymbol{w}}.$$

As $\boldsymbol{Q}$ has exactly 3 eigenvalues (counted according to multiplicity), we obtain with (2.3) for the set eig($\boldsymbol{Q}$) of eigenvalues of $\boldsymbol{Q}$:

$$\text{eig}(\boldsymbol{Q}) = \{e^{ix}, e^{-ix}, 1\} \quad \text{or} \quad \text{eig}(\boldsymbol{Q}) = \{e^{ix}, e^{-ix}, -1\}$$

for some $x \in [0, \pi]$. But for $\text{eig}(\boldsymbol{Q}) = \{e^{ix}, e^{-ix}, -1\}$ we get[4] $\det(Q) = -1$, so that we must have

$$\text{eig}(\boldsymbol{Q}) \; = \; \{e^{ix}, e^{-ix}, 1\} \qquad \text{for } x \in [0, \pi]. \tag{2.4}$$

This also shows that $\lambda = 1$ cannot have multiplicity 2, but only 1 or 3. If $\lambda = 1$ has multiplicity 3, then $\boldsymbol{Q} = \text{Id}$, which therefore is the only case where there can be a non-unique axis of rotation.

For any eigenvalue λ of $\boldsymbol{Q}$, the inverse λ^{-1} will be an eigenvalue of $Q^{-1} = Q^T$ corresponding to the same eigenvector. Therefore, $\boldsymbol{Qw} = \boldsymbol{w}$ immediately implies that $\boldsymbol{w}$ is an element of the nullspace of $\boldsymbol{S} = (\boldsymbol{Q} - \boldsymbol{Q}^T)$, i.e. $\boldsymbol{Sw} = 0$. On the other hand if $\boldsymbol{z}$ is the unique axial vector of the skew matrix $\boldsymbol{S}$, then

$$0 = \boldsymbol{Sw} = \boldsymbol{z} \times \boldsymbol{w}, \tag{2.5}$$

so that $\boldsymbol{z}$ and $\boldsymbol{w}$ are parallel, i.e. the axial vector of the skew matrix is parallel to the axis of rotation of $\boldsymbol{Q}$.

(c) If $\boldsymbol{v}$ is any vector orthogonal to the axis of rotation $\boldsymbol{w}$, then

$$\begin{aligned} \boldsymbol{Qv} \cdot \boldsymbol{w} &= \boldsymbol{v} \cdot \boldsymbol{Q}^T \boldsymbol{w} \\ &= \boldsymbol{v} \cdot \boldsymbol{w} \\ &= 0 \end{aligned} \tag{2.6}$$

Furthermore from (2.1) we can conclude that if $\boldsymbol{v}$ is a unit vector, then so also is $\boldsymbol{Qv}$. Similarly if $\boldsymbol{u}$ is a second unit vector orthogonal to both $\boldsymbol{w}$ and $\boldsymbol{v}$, then the subspace spanned by $\{\boldsymbol{u}, \boldsymbol{v}\}$ is invariant under the action of $\boldsymbol{Q}$, that is,

$$\text{span}\{\boldsymbol{u}, \boldsymbol{v}\} \; = \; \boldsymbol{Q}(\text{span}\{\boldsymbol{u}, \boldsymbol{v}\}).$$

(This follows together with the above considerations from $\boldsymbol{Qu} \cdot \boldsymbol{Qv} = \boldsymbol{u} \cdot \boldsymbol{v} = 0$.)

[4] Here, we make use of

$$\det(A) = \prod_{i=1}^{n} \lambda_i,$$

for any matrix $A \in \mathbb{C}^{n \times n}$ with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ repeated according to (algebraic) multiplicity.

According to (2.4), the eigenvalues of $\boldsymbol{Q}$ are of the form 1 and a complex conjugate pair $e^{\pm ix}$ for some $x \in [0, \pi]$. We now show that x equals the angle of rotation θ between $\boldsymbol{v}$ and $\boldsymbol{Qv}$, when $\boldsymbol{v}$ is a unit vector orthogonal to $\boldsymbol{w}$. As is well-known, the angle θ between $\boldsymbol{v}$ and $\boldsymbol{Qv}$ satisfies the relation:

$$\cos\theta \; = \; \frac{\boldsymbol{Qv} \cdot \boldsymbol{v}}{\|\boldsymbol{Qv}\| \|\boldsymbol{v}\|} \; = \; \boldsymbol{Qv} \cdot \boldsymbol{v}.$$

The same will hold for the angle θ between $\boldsymbol{u}$ and $\boldsymbol{Qu}$ for any unit vector $\boldsymbol{u}$ being orthogonal to $\boldsymbol{w}$ and $\boldsymbol{v}$, such that we get:

$$2 \cdot \cos\theta \; = \; \boldsymbol{Qv} \cdot \boldsymbol{v} \; + \; \boldsymbol{Qu} \cdot \boldsymbol{u}.$$

On the other hand, it is shown at the end that we obtain by using the spectral decomposition together with Parseval's equation the following relation:

$$\boldsymbol{Qv} \cdot \boldsymbol{v} \; + \; \boldsymbol{Qu} \cdot \boldsymbol{u} \; = \; e^{ix} + e^{-ix} \; = \; 2 \cdot \cos x, \tag{2.7}$$

from which we immediately arrive at $x = \theta$.

As the trace of a matrix equals the sum of the eigenvalues (repeated according to algebraic multiplicity), we get:

$$\mathrm{tr}(\boldsymbol{Q}) = 1 + e^{i\theta} + e^{-i\theta} = 1 + 2\cos\theta. \tag{2.8}$$

But the action of any unit complex eigenvalue in its two dimensional eigenspace is a rotation. So

$$\boldsymbol{Q}[\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{u}] = [\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{u}]\boldsymbol{R}, \tag{2.9}$$

where

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}. \tag{2.10}$$

The identity (2.9) is seen by the following consideration: Suppose for the moment that

$$\boldsymbol{w} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} =: \boldsymbol{e}_1, \quad \boldsymbol{v} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} =: \boldsymbol{e}_2, \quad \boldsymbol{u} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} =: \boldsymbol{e}_3.$$

Then it can be easily seen that $\boldsymbol{Q} = \boldsymbol{R}$. However, as we cannot assume the above equalities, the matrix $\boldsymbol{R}$ will only represent the linear mapping $\boldsymbol{Q}$ with respect to the basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$. Therefore, we have to define the basis transformation $\boldsymbol{S}$ with

$$\boldsymbol{S}[\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{u}] \;=\; \mathrm{Id},$$

such that we conclusively obtain

$$\boldsymbol{Q} \;=\; \boldsymbol{S}^{-1}\boldsymbol{R}\boldsymbol{S}.$$

But $\boldsymbol{S}^{-1}$ is given by

$$\boldsymbol{S}^{-1} = [\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{u}],$$

which immediately provides us with (2.9).

One can go further, and write the rotation $\boldsymbol{Q}$ in terms of the unit axial vector $\boldsymbol{w}$ using vector notation. The transformation (2.9) is equivalent to multiplying by the matrix

$$\boldsymbol{Q} = \cos\theta\,\mathrm{Id} + (1 - \cos\theta)\boldsymbol{w} \otimes \boldsymbol{w} + \sin\theta \boldsymbol{w}^{\times}, \tag{2.11}$$

where $\boldsymbol{w}^{\times}$ denotes the skew matrix with $\boldsymbol{w}$ as axial vector (that is $\boldsymbol{w}^{\times} = \boldsymbol{S}$) and $\boldsymbol{w} \otimes \boldsymbol{w}$ is the rank-one outer product matrix. This is verified by computing the RHS of equation (2.9) and using the fact that $-\boldsymbol{u} = \boldsymbol{w} \times \boldsymbol{v}$, $\boldsymbol{v} = \boldsymbol{w} \times \boldsymbol{u}$, $\boldsymbol{w} \otimes \boldsymbol{w}\, \boldsymbol{u} = \boldsymbol{w} \otimes \boldsymbol{w}\, \boldsymbol{v} = 0$, and $\boldsymbol{w} \otimes \boldsymbol{w}\, \boldsymbol{w} = \boldsymbol{w}$. Then if one introduces the Cayley vector η defined by $\eta = |\eta|\boldsymbol{w}$ with $|\eta|$ defined by

$$\frac{2|\eta|}{1 + |\eta|^2} = \sin\theta, \qquad \frac{1 - |\eta|^2}{1 + |\eta|^2} = \cos\theta, \tag{2.12}$$

one finds that

$$\boldsymbol{Q} = (1 - \alpha|\eta|^2)\mathrm{Id} + \alpha\eta \otimes \eta + \alpha\eta^{\times}, \tag{2.13}$$

where

$$\alpha = \frac{2}{1 + |\eta|^2}. \tag{2.14}$$

The Cayley vector is well-defined except for rotations through $\theta = \pi$. Finally (2.13) is equivalent to

$$\boldsymbol{Q} = (\mathrm{Id} + \eta^{\times})(\mathrm{Id} - \eta^{\times})^{-1}, \tag{2.15}$$

where one observes that the matrix $(\mathrm{Id} - \eta^{\times})$ is always invertible because it is the sum of a positive and a skew-symmetric matrix.

Finally, we proof the equality (2.7). To this end, we need the spectral decomposition

$$\boldsymbol{Q} = \boldsymbol{P}\,\boldsymbol{D}\,\boldsymbol{P}^{-1}, \qquad \boldsymbol{D} = (d_{ij})_{i,j} = \operatorname{diag}\{1, e^{ix}, e^{-ix}\},$$

and

$$\boldsymbol{P} = (p_{ij})_{i,j} = [\boldsymbol{w}, \boldsymbol{w}_1, \boldsymbol{w}_2],$$

where

$$\boldsymbol{Q}\boldsymbol{w}_1 = e^{ix}\boldsymbol{w}_1, \quad \boldsymbol{Q}\boldsymbol{w}_2 = e^{-ix}\boldsymbol{w}_2, \quad \|\boldsymbol{w}_1\| = \|\boldsymbol{w}_2\| = 1.$$

Moreover, we will make use of Parseval's equation:

$$\|\boldsymbol{z}\|^2 = \sum_{e \in S} |\boldsymbol{z} \cdot \boldsymbol{e}|^2, \qquad \text{for all } \boldsymbol{z},$$

where S is a basis of orthonormal vectors[5]. Now, recall that $\boldsymbol{u} = (u_1, u_2, u_3)$, $\boldsymbol{v} = (v_1, v_2, v_3)$ are chosen to be unity vectors such that $S = \{\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{u}\}$ is an orthonormal basis of $\mathbb{C}^3$. Exploiting[6] $\boldsymbol{P}^{-1} = \overline{\boldsymbol{P}}^T$ and $\boldsymbol{P}\boldsymbol{z}_1 \cdot \boldsymbol{z}_2 = \boldsymbol{z}_1 \cdot \overline{\boldsymbol{P}}^T \boldsymbol{z}_2$, we obtain

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{v} \cdot \boldsymbol{v} + \boldsymbol{Q}\boldsymbol{u} \cdot \boldsymbol{u} &= \boldsymbol{D}\overline{\boldsymbol{P}}^T\boldsymbol{v} \cdot \overline{\boldsymbol{P}}^T\boldsymbol{v} + \boldsymbol{D}\overline{\boldsymbol{P}}^T\boldsymbol{u} \cdot \overline{\boldsymbol{P}}^T\boldsymbol{u} \\
&= \sum_{i=1}^{3} d_{ii} \left(\overline{\boldsymbol{P}}^T\boldsymbol{v}\right)_i \overline{\left(\overline{\boldsymbol{P}}^T\boldsymbol{v}\right)_i} + \sum_{i=1}^{3} d_{ii} \left(\overline{\boldsymbol{P}}^T\boldsymbol{u}\right)_i \overline{\left(\overline{\boldsymbol{P}}^T\boldsymbol{u}\right)_i} \\
&= \sum_{i=1}^{3} d_{ii} \Big|\sum_{j=1}^{3} \overline{p_{ji}}\, v_j\Big|^2 + \sum_{i=1}^{3} d_{ii} \Big|\sum_{j=1}^{3} \overline{p_{ji}}\, u_j\Big|^2 \\
&= e^{ix}|\overline{\boldsymbol{w}_1} \cdot \boldsymbol{v}|^2 + e^{-ix}|\overline{\boldsymbol{w}_2} \cdot \boldsymbol{v}|^2 + e^{ix}|\overline{\boldsymbol{w}_1} \cdot \boldsymbol{u}|^2 + e^{-ix}|\overline{\boldsymbol{w}_2} \cdot \boldsymbol{u}|^2 \\
&= e^{ix}(|\overline{\boldsymbol{w}_1} \cdot \boldsymbol{v}|^2 + |\overline{\boldsymbol{w}_1} \cdot \boldsymbol{u}|^2) + e^{-ix}(|\overline{\boldsymbol{w}_2} \cdot \boldsymbol{v}|^2 + |\overline{\boldsymbol{w}_2} \cdot \boldsymbol{u}|^2).
\end{aligned}$$

Now we make use of Parselval's equation. To this end, we use $\boldsymbol{w}\cdot\boldsymbol{v} = \boldsymbol{w}\cdot\boldsymbol{u} = 0$ and get

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{v} \cdot \boldsymbol{v} + \boldsymbol{Q}\boldsymbol{u} \cdot \boldsymbol{u} &= e^{ix}\|\overline{\boldsymbol{w}_1}\|^2 + e^{-ix}\|\overline{\boldsymbol{w}_2}\|^2 \\
&= e^{ix} + e^{-ix}.
\end{aligned}$$

---

[5] This can be proven easily by using

$$\boldsymbol{z} = \sum_{e \in S} (\boldsymbol{z} \cdot \boldsymbol{e})\, \boldsymbol{e}, \qquad \Big\|\sum_{e \in \hat{S}} \boldsymbol{e}\Big\|^2 = \sum_{e \in \hat{S}} \|\boldsymbol{e}\|^2,$$

for any orthonormal basis S and any finite subset $\hat{S} \subset S$ of S. The notation is adapted in order to include infinite-dimensional Hilbert spaces (=complete vector spaces with scalar product).

[6] $\boldsymbol{C} = \overline{\boldsymbol{A}}$ means that $\boldsymbol{C}$ contains the complex conjugate components from $\boldsymbol{A}$.

Note that in general $\|\overline{\boldsymbol{z}}\| = \|\boldsymbol{z}\|$. In particular, we have $\overline{\boldsymbol{w}_i} = \boldsymbol{w}_j$ for $i, j = 1, 2$ and $i \neq j$.

3 Quaternion parametrization

(a) Using the expression for $\boldsymbol{Q}$ in terms of the components of the quaternion $\boldsymbol{q}$ and simply computing $\boldsymbol{Qk}$, immediately shows $\boldsymbol{Qk} = \boldsymbol{k}$. If $\boldsymbol{k} = 0$, again using the expression of $\boldsymbol{Q}$ and the fact that $\boldsymbol{q} \cdot \boldsymbol{q} = q_4^2 = 1$, we can conclude that $\boldsymbol{Q}$ is the identity matrix.

(b) Using the result, valid for a general proper rotation matrix $\boldsymbol{Q}$, that

$$\mathrm{tr}(\boldsymbol{Q}) = 1 + 2\cos\theta, \tag{3.1}$$

and also computing the Trace as the sum of the diagonal entries in the expression for $\boldsymbol{Q}$ in the quaterion parametrization, we get

$$4q_4^2 - 1 = 1 + 2\cos\theta \tag{3.2}$$

so that

$$\begin{aligned} 4q_4^2 &= 2(1 + \cos\theta) \\ &= 2(1 + \cos^2\theta/2 - \sin^2\theta/2) \\ &= 4\cos^2\theta/2. \end{aligned} \tag{3.3}$$

(c) Directly computing $\boldsymbol{B}_1\boldsymbol{q}$, $\boldsymbol{B}_2\boldsymbol{q}$, and $\boldsymbol{B}_3\boldsymbol{q}$, shows that the resulting vectors are of unit norm given that $\boldsymbol{q}$ has unit norm. It is also easy to check that the vectors $\boldsymbol{q}$, $\boldsymbol{B}_1\boldsymbol{q}$, $\boldsymbol{B}_2\boldsymbol{q}$, and $\boldsymbol{B}_3\boldsymbol{q}$ are pairwise orthogonal. This orthogonality guarantees the fact that the four vectors are linearly independent. Therefore we conclude that the set is an orthonormal basis for $\mathbb{R}^4$.