2. Some Applications to Optimisation Problems and to Matrices

Optimisation of Convex Functions

Recall that a function $\phi : K \to \mathbb{R}$ – where $K$ is a convex subset of some vector space – is said to be a convex function if

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y) \quad \forall x, y \in K, t \in [0,1]$$

This may be thought of as the geometric requirement that given any two points on the graph, the chord joining them always lies above (or on) the graph; for instance, the function $\mathbb{R} \ni x \mapsto x^k \in \mathbb{R}$ is convex if and only if $k$ is even; likewise, the function $x \mapsto |x|$ is also a convex function.

Linear functions are trivially convex; it follows that if $f : \mathbb{R}^n \to \mathbb{R}$ is a linear function, and if $K \subset \mathbb{R}^n$ is convex, then the equation $\phi(x) = |f(x)|$ defines a convex function.

Thus one consequence of the Krein–Milman theorem is to ‘optimisation problems’, via the following corollary.

**Corollary 1** If $\phi : K \to \mathbb{R}$ is a convex function defined on a compact convex set $K$, then there exists an extreme point $v \in K$ such that $\phi(x) \leq \phi(v) \quad \forall x \in K$

**Proof:** We need to show that $\max\{\phi(x) : x \in K\} = \max\{\phi(v) : v \in \partial_e(K)\}$. If we denote the maximum on the left side (resp., right side) of the last equation by $M$ (resp., $M_e$), it is clear that $M \geq M_e$. So we need to

---

show that
\[ z \in K \Rightarrow \phi(z) \leq M_e \] (1)

We prove this in stages:

Case (i): \( z \in \partial_e(K) \). This is obvious.

Case (ii): \( z \in \text{conv}(\partial_eK) \).

If \( z = \sum_{i=1}^{m} \theta_i v_i \) is an expression of \( z \) as a convex combination of extreme points, then
\[
\phi(z) \leq \sum_{i=1}^{m} \theta_i \phi(v_i) \leq \sum_{i=1}^{m} \theta_i M_e = M_e
\]

Case (iii): \( z \in K \) arbitrary

Then, the Krein–Milman theorem allows us to choose a sequence \( \{z_k\}_k \subset \text{conv}(\partial_e(K)) \) such that \( ||z_k - z|| \to 0 \); since \( \phi(z_k) \leq M_e \ \forall k \), we find\(^2\) that also \( \phi(z) \leq M_e \), as desired.

In the typical ‘linear programming problem’, one wishes to maximise some linear function \( L \) of \( n \) variables subject to some ‘linear constraints’, i.e., conditions of the form \( f_i(x_1, \ldots, x_n) \leq C_i, \ 1 \leq i \leq N \) where the \( f_i \) are linear functions. Typically this comes from a problem arising ‘from the market’, and the set \( K = \{ x \in \mathbb{R}^n : f_i(x) \leq C_i \ \forall 1 \leq i \leq N \} \) is compact and convex, and the function \( L \) to be optimised need then only be optimised over the smaller set \( \partial_e(K) \). In practice, the set \( K \) is a ‘polytope’ or the set of extreme points is finite, and one only then needs an efficient algorithm – such as the ‘simplex method’ or refinements of the sort that made Karmarkar famous – to scan all the values of \( L \) over this finite set in order to identify the maximum value of \( L \).

Applications to Matrices

One of the most important problems that needs to be solved in a variety of applications is that of determining...
the eigenvalues of a matrix. Very often, even the matrix (whose eigenvalues are sought) is known only approximately. (Think of ‘feeding’ a matrix, which has some irrational numbers as entries, to a computer!) We would like to know, for instance, that if we knew the entries of a matrix to within some prescribed ‘error’ which we can control, then the ‘error’ in the computed eigenvalue is within acceptable standards. We shall try to indicate how this problem is solved as a result of three results – namely, the Krein–Milman theorem, the Birkhoff–von Neumann theorem, and the Hoffmann–Wielandt theorem; but first, we go through a quick ‘crash course’ in the pertinent linear algebra.

We will need first to understand the relationship between linear transformations and matrices. Recall that a mapping \( T : \mathbb{R}^n \to \mathbb{R}^n \) is said to be a linear transformation if it satisfies

\[
T(sx + ty) = sTx + tTy \quad \forall \ s, t \in \mathbb{R}, \ x, y \in \mathbb{R}^n
\]

Recall that an (ordered) orthonormal basis for \( \mathbb{R}^n \) is a collection \( B = \{e_1, \ldots, e_n\} \) in \( \mathbb{R}^n \) such that

\[
\langle e_i, e_j \rangle = \delta_{ij} = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{otherwise}
\end{cases}
\]

If \( B \) is as above, then it is true that \( x = \sum_i (x, e_i)e_i \) for all \( x \in \mathbb{R}^n \). Define \( [T]_B \) to be the matrix (or array) given by

\[
[T]_B = A = ((a_{ij})) = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\]

where \( a_{ij} = (Te_j, e_i) \); thus, if we make the identification

\[
z \leftrightarrow \begin{bmatrix}
(z, e_1) \\
\vdots \\
(z, e_n)
\end{bmatrix}
\]
we see that the matrix $[T]_B$ is defined by the requirement that its $j$-th column is precisely the vector $T e_j$ (in the above correspondence). Thus, if $T$ and $A$ are related as above, then

$$T e_j = \sum_i a_{ij} e_i \quad \text{(2)}$$

Conversely, if $A$ is an arbitrary $n \times n$ matrix, and if we define the transformation $T$ by

$$T x = \sum_j (x, e_j) T e_j = \sum_i (\sum_j a_{ij} (x, e_j)) e_i$$

it is easily verified that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation such that $[T]_B = A$.

Thus, we see that once we choose and fix some (ordered) orthonormal basis $B$ for $\mathbb{R}^n$ then we have a 1-1 correspondence

$$\mathcal{L}(\mathbb{R}^n) \ni T \leftrightarrow [T]_B \in M_n(\mathbb{R}) \quad \text{(3)}$$

between the collection $\mathcal{L}(\mathbb{R}^n)$ of all linear transformations on $\mathbb{R}^n$ and the collection $M_n(\mathbb{R})$ of all $n \times n$ (real) matrices. There is also a natural bijection between $M_n(\mathbb{R})$ and $\mathbb{R}^{n^2}$, and we may hence ‘transport’ various notions from $\mathbb{R}^{n^2}$ and make sense of such things as (entry-wise) addition of matrices, scalar multiplication of a matrix by a real number ($\alpha \cdot ((a_{ij})) = ((\alpha a_{ij}))$), and the (so-called ‘Hilbert–Schmidt’ or ‘Frobenius’) norm of a matrix (given by $\|((a_{ij}))\| = \left(\sum_{i,j=1}^{n} a_{ij}^2\right)^{\frac{1}{2}}$).

Finally, the ‘composite of two linear transformations’ is clearly again a linear transformation; and hence there exists a unique way to define products of matrices in such a way that the linear transformation corresponding to the ‘matrix-product’ is precisely the composition of the linear transformations associated to the factors. This product is readily verified to be the usual product defined thus: $C = AB$, where $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. 
Remark 2. We wish to single out some special classes of linear transformations and the corresponding matrices.

(1) A linear transformation $T \in \mathcal{L}(\mathbb{R}^n)$ is said to be an isometry if it satisfies the following equivalent conditions:

(i) $T$ maps every (equivalently, any) orthonormal basis of $\mathbb{R}^n$ into an orthonormal basis.

(ii) $(Tx, Ty) = (x, y)$ for all $x, y \in \mathbb{R}^n$

It is seen from (i) that a matrix corresponds to an isometry precisely when its columns form an orthonormal basis. Such matrices are called orthogonal matrices. These may be equivalently described by the requirement that the matrix is invertible and its inverse is its transposed matrix: i.e., if $U^t$ is defined to be the matrix whose $(i, j)$-th entry is $u_{ji}$, then $U$ is an orthogonal matrix precisely when $U^tU$ is the identity matrix $I_n$.

Isometries – the linear distance-preserving maps of $\mathbb{R}^n$ – come up naturally when we wish to understand the relationship between the matrices $A = [T]_B$ and $A' = [T]_{B'}$ of one linear transformation with respect to two orthonormal bases, say, $B = \{e_i : 1 \leq i \leq n\}$ and $B' = \{e'_i : 1 \leq i \leq n\}$, of $\mathbb{R}^n$. It is then true that the unique linear map satisfying $W(e_i) = e'_i \forall i$ is an isometry and that $A = U A' U^t$ where $U = [W]_B$ (is an orthogonal matrix). Thus, two matrices $A$ and $A'$ ‘represent the same transformation’ with respect to two orthonormal bases if and only if there exists an orthogonal matrix $U$ such that $A = U A' U^t$.

(2) A linear transformation $A \in \mathcal{L}(\mathbb{R}^n)$ is said to be self-adjoint if $(Tx, y) = (x, Ty) \forall x, y \in \mathbb{R}^n$; we see that a matrix corresponds to a self-adjoint transformation precisely when it is symmetric; i.e., $a_{ij} = a_{ji} \forall i, j$.

The content of the ‘spectral’ or ‘principal axes’ theorem
is that self-adjoint transformations are diagonalisable – a transformation $T$ being said to be diagonalisable if there exists an orthonormal basis $\{B = e_i\}$ and scalars $\alpha_i$ such that $T e_i = \alpha_i e_i$ for all $i$; thus, the matrix of $T$ with respect to $B$ is the diagonal matrix

$$A = [T]_B = \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ \vdots & \vdots & \ddots \\ 0 & 0 & \alpha_n \end{bmatrix} = D(\alpha)$$

The numbers $\alpha_j$ are called the eigenvalues of $T$ (resp., $A$).

In view of the remarks in (1) above, we see thus that if $B$ is any symmetric matrix, then there exists an orthogonal matrix $U$ and $B = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$ such that $B = UD(\beta)U^t$, where, of course, $D(\beta)$ denotes the diagonal matrix with $(i, i)$-th coordinate $\beta_i$.

Recall that an $n \times n$ matrix $P$ is said to be doubly stochastic if it satisfies the following conditions:

(i) $p_{ij} \geq 0 \ \forall \ 1 \leq i, j \leq n$;

(ii) $\sum_{i=1}^{n} p_{ij} = \sum_{j=1}^{n} p_{ij} = 1$.

The collection $\mathcal{D}_n$ of all such doubly stochastic matrices is easily seen to be a convex subset of (the vector space) $M_n(\mathbb{R})$, which is closed and bounded, i.e., compact (where we think of $M_n(\mathbb{R})$ as an ‘isomorphic’ copy of $\mathbb{R}^{n^2}$). Verify this!

An example of a doubly stochastic matrix is the identity matrix (which has 1’s on the main diagonal and 0’s elsewhere). A less trivial example of an element of $\mathcal{D}_3$ is given by

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

More generally, recall that a permutation of $\{1, 2, \ldots, n\}$
is a 1-1 function of \(\{1, 2, \ldots, n\}\) onto itself; given such a permutation, call it \(\sigma\), the associated permutation matrix is the matrix \(P(\sigma) \in M_n(\mathbb{R})\) defined by

\[
p_{ij}^{(\sigma)} = \begin{cases} 
1 & \text{if } i = \sigma(j) \\
0 & \text{otherwise}
\end{cases}
\]

That we even have \(P(\sigma) \in \mathcal{D}_n\) is the 'easy half' of the following result.

**Theorem 3 (Birkhoff–von Neumann)**

The extreme points of the set \(\mathcal{D}_n\) are precisely the permutation matrices.

We do not prove this theorem here; the proof is not too difficult; the interested reader should try to furnish an independent proof. We introduce some notation and list some simple facts as exercises.

**Exercise** Denote the collection of all permutations of \(\{1, 2, \ldots, n\}\) by \(S_n\), and as before, write \(P(\sigma)\) to denote the permutation matrix associated to \(\sigma \in S_n\). Verify the following assertions:

1. \(S_n\) is a group with respect to composition as product (i.e., if \(\sigma, \tau \in S_n\), then \((\sigma \tau)(i) = \sigma(\tau(i))\)) and the identity element of this group is the identity permutation \(1(i) = i \ \forall i\).

2. If \(\sigma \in S_n\), then \(P(\sigma^{-1}) = (P(\sigma))^{-1} = (P(\sigma))^t\), and consequently each permutation matrix is an orthogonal matrix.

3. The map \(\sigma \mapsto P(\sigma)\) defines a homomorphism of the group \(S_n\) into the group of orthogonal matrices – i.e., \(P(\sigma)P(\tau) = P(\sigma \tau) \ \forall \sigma, \tau \in S_n\).

The norm and inner product used in \(M_n(\mathbb{R})\) in the following result (and its proof) are the usual ones obtained by identifying \(M_n(\mathbb{R})\) with \(\mathbb{R}^{n^2}\); thus, if \(A = ((a_{ij})), B = ((b_{ij}))\), then \((A, B) = \sum_{ij} a_{ij}b_{ij}\).
We shall need the fact that if $U$ is an orthogonal matrix and if $A \in M_n(\mathbb{R})$ is arbitrary, then $\|UA\| = \|A\| = \|AU\| = \|A^t\|$. (The proof is not hard, and the reader is urged to try and prove it for himself; if necessary, he may find a proof of this assertion in [1], for instance.)

**Theorem 4 (Hoffmann–Wielandt)**

Suppose $A, B \in M_n(\mathbb{R})$ are symmetric matrices, with eigenvalues $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$, respectively. Let us write $\alpha = (\alpha_1, \ldots, \alpha_n), \beta = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$.

Then,

$$\|A - B\| \geq \min_{\sigma \in S_n} \|\alpha - P^{(\sigma)}\beta\|$$

**Proof:** We may, in view of Remark (2), assume without loss of generality that $A = D(\alpha)$ and $B = UD(\beta)U^t$, where $U$ is some isometry. Then, compute as follows:

$$\|A - B\|^2 = \|D(\alpha) - UD(\beta)U^t\|^2$$

$$= \|D(\alpha)\|^2 + \|UD(\beta)U^t\|^2 - 2(D(\alpha), UD(\beta)U^t)$$

$$= \|D(\alpha)\|^2 + \|D(\beta)\|^2 - 2(D(\alpha), UD(\beta)U^t)$$

Notice now that

$$(D(\alpha), UD(\beta)U^t) = \sum_i \alpha_i (UD(\beta)U^t)_{ii}$$

$$= \sum_i \alpha_i \sum_j (u_{ij})^2 \beta_j$$

writing $p_{ij} = (u_{ij})^2$, we see (from the orthogonality of $U$) that $P = ((p_{ij})) \in D_n$, and we may, thanks to the Birkhoff–von Neumann theorem and the Krein–Milman theorem, write $P = \sum_{\sigma \in S_n} \theta_\sigma P^{(\sigma)}$, where $\theta_\sigma \geq 0 \forall \sigma$ and $\sum_{\sigma \in S_n} \theta_\sigma = 1$. Hence we find that

$$\|A - B\|^2 = \|D(\alpha)\|^2 + \|D(\beta)\|^2 - 2 \sum_i \alpha_i (P\beta)_i$$

$$= \|D(\alpha)\|^2 + \|D(\beta)\|^2 - 2(\alpha, P\beta)$$

$$= \|\alpha\|^2 + \|\beta\|^2 - 2 \sum_{\sigma \in S_n} \theta_\sigma (\alpha, P^{(\sigma)}\beta)$$
\[
\sum_{\sigma \in S_n} \theta_{\sigma} \left( \|\alpha\|^2 + \|\beta\|^2 - 2(\alpha, P^{(\sigma)}\beta) \right) \\
= \sum_{\sigma \in S_n} \theta_{\sigma} \|\alpha - P^{(\sigma)}\beta\|^2 \\
\geq \min_{\sigma \in S_n} \|\alpha - P^{(\sigma)}\beta\|^2
\]

as desired. \[\Box\]

The reader is urged to convince himself that the Hoffmann-Wielandt theorem does indeed solve the problem stated in the first paragraph of this subsection.

**Conclusion**

In conclusion, we find that the Krein-Milman theorem is a simply stated and geometrically pleasing fact, which has manifold applications. It is also a powerful theoretical result, and to see its full strength, it should be stated as a result about compact convex sets in a general (possibly infinite-dimensional) *locally convex topological space*. The interested reader is directed to such standard texts as [2] for further details.

**Suggested Reading**


---

*I am not qualified to talk about him as a teacher and a scientist. He was committed to this side of his life with his greatest passion and intensity. With smiling certainty, he once said to me: “I was lucky enough to be allowed once to look over the good Lord’s shoulder while He was at work: “ That was enough for him, more than enough! It gave him great joy, and the strength to meet the hostilities and misunderstandings he was subjected to in the world time and again with equanimity, and not to be led astray.*

*Elisabeth Heisenberg on Werner Heisenberg*