NOTES FOR 5 OCT (TUESDAY)

1. Recap

(1) Defined Hermitian operators and proved the uncertainty principle.
(2) Defined orthonormal bases and did Grahm-Schmidt. Used it to prove the Cholesky decomposition.

2. Orthogonal projections

The Gram-Schmidt procedure is a special case of an operation called the orthogonal projection. To develop this concept, firstly, here is a natural definition : Let $v \in V$ and let $W \subset V$ be a subspace. A vector $w \in W$ is said to be a best approximation of v from W if it minimises $\|w - v\|$. It is not hard to have a geometric picture of the best approximation. However, in infinite-dimensions it gets a little tricky. Nonetheless, here is a general theorem.

Theorem 2.1. *Let W be a subspace of an inner product space V, and let $v \in V$.*

(1) *A vector $w \in W$ is a best approximation of v from W iff $v - w$ is orthogonal to every vector in W.*
(2) *If a best approximation exists, it is unique.*
(3) *If W is finite-dimensional, and $e_1, \ldots, e_n$ is any orthonormal basis for W, then $w = \sum_i \langle v, e_i \rangle e_i$ is the best approximation of v from W.*

Proof. (1) If w is a best approximation, then given a vector $\alpha \in W$, the function $f(t) = \|v - w + t\alpha\|^2$ achieves a local minimum at $t = 0$. So $f'(0) = 0 = 2Re(\langle v - w, \alpha \rangle)$. Likewise, replacing α by $\sqrt{-1}\alpha$, we see that $0 = 2Im(\langle v - w, \alpha \rangle)$. Thus, $v - w$ is orthogonal to every such vector. Conversely, if $v - w$ is orthogonal to every vector and if $w' \in W$, then $w' - w \in W$ and hence it is easy to see that $\|v - w'\|^2 = \|v - w\|^2 + \|w' - w\|^2 \geq \|v - w\|^2$.
(2) If w, w' are two best approximations, then by the previous part, $w = w'$.
(3) By a direct calculation, $v - w$ is orthogonal to e_i for all i and hence orthogonal to all vectors in W. We are done.

□

Here is a related definition : If $S \subset V$ is a subset, the orthogonal complement $S^\perp$ is the set of vectors in V that are orthogonal to S. Note that it is always a subspace. (Why?) Also, $\{0\}^\perp = V$ and $V^\perp = \{0\}$. When the best approximation $w \in W$ to v exists, then w is called the orthogonal projection of v on W. If every vector has an orthogonal projection to W, then the map $\Pi_W : V \to W$ taking v to its orthogonal projection is called the orthogonal projection map to W.
Firstly, Π_W is a linear map : Indeed, $\langle au + bv - (a\Pi_W u + b\Pi_W v), w \rangle = a\langle u - \Pi_W u, w \rangle + b\langle v - \Pi_W v, w \rangle = 0$. Hence, $\Pi_W(au + bv) = a\Pi_W u + b\Pi_W v$. Secondly, $\Pi_W : W \to W$ is the identity. Thirdly, $\Pi_W^2 = I \circ \Pi_W = \Pi_W$. (So if Π_W exists, it is "idempotent".) Fourthly, if Π_W exists, then $I - \Pi_W$ is the orthogonal projection to $W^\perp$. Indeed, $\langle v - \Pi_W v, w \rangle = 0$ for all $w \in W$ and hence it is a linear map to $W^\perp$. Moreover, $\langle v - (v - \Pi_W v), w' \rangle = \langle \Pi_W v, w' \rangle = 0$ for all $w' \in W^\perp$. Hence, $v - \Pi_W v$ is the best approximation to v from $W^\perp$. So, $\Pi_{W^\perp} = I - \Pi_W$. Fifthly, if Π_W exists, then $\Pi_W v = 0$ iff $\langle v, w \rangle = 0$ for all $w \in W$ and hence iff $v \in W^\perp$. Lastly, if Π_W exists, then $W \oplus W^\perp \equiv V$. Indeed, define $T(w, w') = w + w'$. If $T(w, w') = 0$, then $w = -w' \in W \cap W^\perp$, i.e., $\langle w, w \rangle = 0 = \langle w', w' \rangle$. Thus $w = w' = 0$. Moreover, $v = \Pi_W v + (I - \Pi_W)v$.

Hence we are done. Note, that if W is finite-dimensional, Π_W necessarily exists. Likewise, if Π_W exists, then $(W^\perp)^\perp = W$. This is because $W^\perp = ker(\Pi_W)$ and $(W^\perp)^\perp = ker(\Pi_{W^\perp}) = ker(I - \Pi_W)$ which consists of v such that $v = \Pi_W v$, i.e., $v \in W$.

So the Grahm-Schmidt process is : Let Π_k be the orthogonal projection onto the orthogonal complement of $w_1, \ldots, w_{k-1}$, and let $\Pi_1 = \frac{1}{\|w_1\|} I$. Then, defining $v_k = \frac{\Pi_k w_k}{\|\Pi_k w_k\|}$ is the Gram-Schmidt process.

Note that the observations imply that if $e_1, \ldots, e_k$ is an orthonormal set of vectors in V, then $\sum_k |\langle v, e_k\rangle|^2 \leq \|v\|^2$. Indeed, $v = (I - \Pi)v + \Pi v$ where Πv is the orthogonal projection to the subspace spanned by e_i. So $\|\Pi v\|^2 \leq \|v\|^2$ with equality iff v is in the subspace. As a consequence, $\sum_{k=-n}^{n} |\int_0^1 f(t) e^{-2\pi\sqrt{-1}kt} dt|^2 \leq \int_0^1 |f(t)|^2 dt$. This inequality is called Bessel's inequality.

There is an interesting way to write an orthogonal projection onto a unit vector w in $\mathbb{C}^n$ with the usual inner product. Indeed, $\Pi_w v = \langle v, w\rangle w = \sum_{i,j} v_j \bar{w}_j w_i e_i$. This motivates the following definition : Given two vectors $u, v \in \mathbb{C}^n$, the matrix $A_{ij} = u_j \bar{v}_i$ is called the outer product of u and v and the corresponding linear map is written (usually by physicisits) as $|u\rangle\langle v|$. Note that $(\Pi_w v)_i = (|w\rangle\langle w|v)_i$. So if $w_1, \ldots, w_k$ is an orthonormal basis for a subspace W, then $\Pi_W = \sum_i |w_i\rangle\langle w_i|$.

3. Linear functionals and adjoints

Given a linear map $T : V \to W$, there exists a natural map $T^* : W^* \to V^*$ given by $T^*(\lambda)(v) = \lambda(Tv)$. This map is called the adjoint map. If V is an inner product space, there is a more interesting formulation of this adjoint map. Indeed, firstly, we have a special case of the so-called "Little Riesz representation theorem".

Lemma 3.1. *If V is a finite-dimensional inner product space, then the map $V \to V^*$ given by $L(v) = \langle , v\rangle$ is a $1-1$ onto map that is "antilinear", i.e., $L(av + bw) = \bar{a}L(v) + \bar{b}L(w)$.*

Proof. Firstly, $L(v)(w) = \langle w, v\rangle$ is linear in w and hence $L(v) \in V^*$. Secondly, it is clearly antilinear in v. Thirdly, if $L(v) = 0$, then $\langle w, v\rangle = 0 \ \forall\ w$. Thus, $\|v\|^2 = 0$ and $v = 0$. So it is $1-1$. If e_i is an orthonormal basis, and $\lambda \in V^*$, then $\lambda(v) = \sum_i v_i \lambda(e_i) = \langle v, \sum_i \overline{\lambda(e_i)} e_i\rangle = L(\sum_i \overline{\lambda(e_i)} e_i)$. So it is onto. □

So V and V^* can be "identified" in a sense for finite-dimensional vector spaces. This means that $T^* : V^* \to V^*$ induces a map (also, confusingly called the adjoint of T) $T : V \to V$ for finite-dimensional vector spaces as follows : There exists a unique linear map $T^* : V \to V$ satisfying $\langle Tv, w\rangle = \langle v, T^* w\rangle$. Indeed, consider the linear functional $\lambda(v) = \langle Tv, w\rangle$. The above theorem implies that there is a unique vector $\tilde{w}_w \in V$ such that $\lambda(v) = \langle v, \tilde{w}_w\rangle$. Noticing that $\langle v, a\tilde{w}_x + b\tilde{w}_y\rangle = \bar{a}\langle v, \tilde{w}_x\rangle + \bar{b}\langle v, \tilde{w}_y\rangle = \bar{a}\langle Tv, x\rangle + \bar{b}\langle Tv, y\rangle = \langle Tv, ax + by\rangle$ and that $\tilde{w}_{ax+by}$ is unique, it is equal to $a\tilde{w}_x + b\tilde{w}_y$. Thus, $T^* w = \tilde{w}_w$ is the unique linear map satisfying the desired relation.