MATHEMATICAL TOOLS - SOLUTION 8

Problem 1.

(1) Let $v_1, \dots, v_n$ and $w_1, \dots, w_m$ be orthogonal bases for V and W, respectively. In order to define a linear map it's enough to define its action on a basis. So define: $T^*(w_i) = \sum_{j=1}^n \langle Tv_j, w_i \rangle_W v_j$. Note that for all i, j: $\langle T^* w_i, v_j \rangle_V = \langle Tv_j, w_i \rangle_W$Then, for all $\sum \alpha_i v_i \in V, \sum \beta_i w_i \in W$:

$$\left\langle T \sum \alpha_i v_i, \sum \beta_j w_j \right\rangle_W = \sum_{i,j} \alpha_i \beta_j \langle Tv_i, w_j \rangle_W = \sum_{i,j} \alpha_i \beta_j \langle T^* w_i, v_j \rangle_V = \left\langle T^* \sum \beta_j w_j, \sum \alpha_i v_i \right\rangle_V$$

as desired.
For the uniqueness, assume S has the desired property. Then for all i, j:

$$\langle Sw_i, v_j \rangle_V = \langle Tv_j, w_i \rangle_W = \langle T^* w_i, v_j \rangle_W$$

so $Sw_i = T^* w_i$ hence $S = T^*$.

(2) Let $\varphi : V \to \mathbb{R}^n$ be the isomorphism given by $v_i \mapsto e_i$ and $\psi : W \to \mathbb{R}^m$ be given by $w_i \mapsto e_i$. One can verify that these isomorphisms preserve the inner product (i.e. $\forall u, v \in V, \langle u, v \rangle_V = \langle \varphi u, \varphi v \rangle$ where $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbb{R}^n$. A similar inequality holds for ψ). Let $S = \psi \circ T \circ \varphi^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ and A be a matrix w.r.t. the standard bases. We claim that A^T represents $\varphi \circ T^* \circ \psi^{-1}$: Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Then

$$\langle x, \varphi T^* \psi^{-1} y \rangle = \langle \varphi^{-1} x, T^* \psi^{-1} y \rangle_V = \langle T \circ \varphi^{-1} x, \psi^{-1} y \rangle_W = \langle \psi^{-1} S x, \psi^{-1} y \rangle_W = \langle Sx, y \rangle = \langle Ax, y \rangle = \langle x, A^T y \rangle$$

So for all x, y $\langle x, \varphi T^* \psi^{-1} y \rangle - \langle x, A^T y \rangle = 0 \implies \forall y \in \mathbb{R}^m, \langle \varphi T^* \psi^{-1} y - A^T y, \varphi T^* \psi^{-1} y - A^T y \rangle = 0 \implies \forall y, \varphi T^* \psi^{-1} y = A^T y$.
Now, from the SVD theorem we know there exist orthogonal bases $a_1, \dots, a_n$ and $b_1, \dots, b_m$ for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, s.t. $Aa_i = \sigma_i b_i$, $A^T b_i = \sigma_i a_i$. Thus for all i, $\varphi T^* \psi^{-1} b_i = \sigma_i a_i \implies T^* \psi^{-1} b_i = \sigma_i \varphi^{-1} a_i$, and similarly $\psi \circ T \circ \varphi^{-1} a_i = \sigma_i b_i \implies T \circ \varphi^{-1} a_i = \sigma_i \psi^{-1} b_i$. So we can take $\{\varphi^{-1} a_i\}$ and $\{\psi^{-1} b_i\}$ as our bases, and we just need to verify that they are orthogonal. Indeed:

$$\langle \varphi^{-1} a_i, \varphi^{-1} a_j \rangle_V = \langle a_i, a_j \rangle = \delta_{i,j}$$

and a similar calculation holds for $\{\psi^{-1} b_i\}$.

Problem 2.

(1) Define $A_{i,j} = \langle v_i, v_j \rangle_1$. Then $\langle \sum \alpha_i v_i, \sum \beta_j v_j \rangle_1 = \sum_{i,j} \alpha_i \beta_j \langle v_i, v_j \rangle_1 = \sum_{i,j} \alpha_i \beta_j A_{i,j}$. But $\alpha_i = \langle \sum \alpha_j v_j, v_i \rangle_V$ and a similar equality holds for the βs, so we're done.

(2) $A_{i,j} = \langle v_i, v_j \rangle_1 = \langle v_j, v_i \rangle_1 = A_{j,i}$ by symmetry of inner products. A is positive definite because:

$$x^T A x = \sum_{i,j} x_i x_j A_{i,j} = \left\langle \sum x_i v_i, \sum x_j v_j \right\rangle_1 \geq 0$$

with equality iff $\sum x_i v_i = 0$, and since $\{v_i\}$ is a basis this happens iff $x = 0$.

(3) Straightforward verification.

Problem 3.

(1) Immediate verification.
(2) $x^T\left(vv^T\right)x = \left(v^Tx\right)^T\left(v^Tx\right) \geq 0$.
(3) Let $A = UDU^T$ be an orthogonal diagonalization of A. Let $v_1, \ldots, v_n$ be the columns of U. Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the diagonal elements of D (these can be assumed to be in descending order). Then $A_{i,j} = \sum_k \lambda_k U_{i,k} U^T_{k,j} = \sum_k \lambda_k U_{i,k} U_{j,k} = \sum_k \lambda_k \left(u_k u_k^T\right)_{i,j}$. It remains to show that $\lambda_i \geq 0$. Well, by positive semi-definiteness:

$$0 \leq \langle Av_i, v_i\rangle = \langle \lambda_i v_i, v_i\rangle = \lambda_i \langle v_i, v_i\rangle = \lambda_i$$

Because v_i is a normalized eigenvector of λ_i.

Problem 4.

(1) Let v be a normalized eigenvector with eigenvalue λ. Then $\lambda = \lambda\langle v, v\rangle = \left\langle A^TAv, v\right\rangle = \langle Av, Av\rangle \geq 0$.
(2) Let v be an eigenvector with eigenvalue λ. Let $u = \frac{1}{\sqrt{\lambda}}Av$. Then $Av = \sqrt{\lambda}u$. $A^TAv = \sqrt{\lambda}A^Tu \implies A^Tu = \frac{1}{\sqrt{\lambda}}A^TAv = \sqrt{\lambda}v$ using the fact that $A^TAv = \lambda v$.
(3) Well, $A^TAv = A^T\sigma u = \sigma^2 v$, and a similar calculation for AA^Tu.

Problem 5.

(1) Assume $A = \begin{pmatrix} x_1 \\ \ldots \\ x_m \end{pmatrix} = x \in \mathbb{R}^m$. Let $v_1 = e_1 = (1), \sigma_1 = \|x\|_2$ and $u_1 = \frac{x}{\|x\|_2}$. Then: $Av_1 = Ae_1 = x = \sigma_1 u_1$ and $A^Tu_1 = x^T\frac{x}{\|x\|_2} = \frac{\|x\|_2^2}{\|x\|_2} = \sigma_1 v_1$. Complete u_1 to an orthonormal basis and we're done.
(2) First, note that for any x, y $f\left(\frac{x}{\|x\|_2}, \frac{y}{\|y\|_2}\right) = \frac{y^TAx}{\|y\|_2\|x\|_2\left\|\frac{x}{\|x\|_2}\right\|_2\left\|\frac{y}{\|y\|_2}\right\|_2} = \frac{y^TAx}{\|x\|_2\|y\|_2} = f(x, y)$. Second note that f attains a maximum on $B^n \times B^m$ as a continuous function on a compact set (this can also be obtained by considering $\|A\|_{op}$ and using Cauchy-Schwarz). Let $(v, u) \in B^n \times B^m$ be a point where f is maximal on $B^n \times B^m$ and let x, y be any two non-zero points. Then, because $\left(\frac{x}{\|x\|_2}, \frac{y}{\|y\|_2}\right) \in B^n \times B^m$:

$$f(x, y) = f\left(\frac{x}{\|x\|_2}, \frac{y}{\|y\|_2}\right) \leq f(v, u)$$

so (v, u) is in fact a global maximum.
(3) f is a continuously differentiable function. By Fermat's theorem, $\nabla f(x, y) = 0$ whenever (x, y) is a local extremum of f. In particular $\nabla f(v, u) = 0$ because (v, u) is a global, hence local, extremum.
(4) $\langle Ax, u_1\rangle = \left\langle x, A^Tu_1\right\rangle = \langle x, \sigma_1 v_1\rangle = \sigma_1\langle x, v_1\rangle = 0$.
(5) Let $a_2, \ldots, a_n \in \mathbb{R}^{n-1}$ and $b_2, \ldots, b_m \in \mathbb{R}^{m-1}$ and $\sigma_2 \geq \ldots \geq \sigma_n$ be as per the inductive hypothesis. Define, for $i \geq 2$: $v_i = \sum_j (a_i)_j x_j$ and $u_i = \sum_j (b_i)_j y_j$. We claim that these are orthonormal, together with v_1, u_1.

First, for $i, j \geq 2$:

$$\langle v_i, v_k \rangle = \left\langle \sum_j (a_i)_j \, x_j, \sum_j (a_k)_j \, x_j \right\rangle = \sum_j (a_i)_j \, (a_k)_j = \langle a_i, a_k \rangle = \delta a_{i,k}$$

and similarly with u. For $i = 1$ recall that $x_2, \ldots, x_n \in \{v_1\}^{\perp}$. Second, recalling that $Tx_i = \sum_j \tilde{A}_{j,i} y_j$:

$$Av_i = A \sum_j (a_i)_j \, x_j = \sum_j (a_i)_j \sum_k \tilde{A}_{k,j} y_k = \sum_k \left(\sum_j (a_i)_j \, \tilde{A}_{k,j} \right) y_k = \sum_k \left(\tilde{A} a_i \right)_k y_k = \sigma_i \sum_k (b_i)_k \, y_k = \sigma_i u_i$$

And a similar calculation holds for $A^T u_i$.