Solutions to problem set #10

ad problem 1: (a) The required SVD is $A^T = V\Sigma^T U^T$.

(*Proof:* In fact, $A = U\Sigma V^T$ leads to $A^T = \left(U\Sigma V^T\right)^T = \underbrace{\left(V^T\right)^T}_{=V} \Sigma^T U^T = V\Sigma^T U^T$, and each of V and U is a matrix with orthonormal columns.)

Don't forget the transposition in Σ^T ! Unless Σ is a **square** matrix, $\Sigma^T \neq \Sigma$ even though Σ is "diagonal". The matrices Σ^T and Σ share the same diagonal entries and their off-diagonal entries are all 0, but their shapes are distinct unless Σ is square.

(b) The required SVD is $A^{-1} = V\Sigma^{-1}U^T$.

(*Proof:* Assume that A is invertible. Then, A is square and has full rank. Hence, U, Σ and V^T must be invertible square matrices (since $A = U\Sigma V^T$), and thus V is an invertible square matrix. Now, $A = U\Sigma V^T$ leads to $A^{-1} = \left(U\Sigma V^T\right)^{-1} = \left(V^T\right)^{-1} \Sigma^{-1} U^{-1}$.

But being a square matrix with orthonormal columns, U must be orthogonal, so that $U^{-1} = U^T$ and similarly $V^{-1} = V^T$. Hence, $A^{-1} = \underbrace{\left(V^T\right)^{-1}}_{\substack{=V \\ \text{(since } V^{-1}=V^T\text{)}}} \Sigma^{-1} \underbrace{U^{-1}}_{=U^T} = V\Sigma^{-1}U^T$.)

ad problem 2: I will only show the results. For how to compute it, see Michael's recitation #10[1] and the solutions to Alex's review problems to exam #3.

(a) Let $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. An SVD is $A = U\Sigma V^T$ with

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad V = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}.$$

(b) Let $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$. An SVD is $A = U\Sigma V^T$ with

$$U = \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

[1] http://web.mit.edu/18.06/www/Fall14/Recitation10_Michael.pdf

(c) Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. An SVD is $A = U\Sigma V^T$ with

$$U = \begin{pmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & 0 \\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$V = \begin{pmatrix} \frac{1}{3}\sqrt{3} & -\frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{3} & \frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{3} & 0 & \frac{1}{3}\sqrt{6} \end{pmatrix}.$$

[There is one possible source of confusion in part **(c)**: here you have to construct orthonormal bases of subspaces which are not always 1-dimensional. Specifically, the eigenspace of $A^T A$ for eigenvalue 0 and the nullspace $N\left(AA^T\right)$ of AA^T are 2-dimensional. You cannot just take an arbitrary basis and scale its vectors to have length 1 here; the resulting vectors might fail to be orthogonal. Instead you need to apply the Gram-Schmidt orthonormalization algorithm.

Let us show this on the example of the eigenspace of $A^T A$ for eigenvalue 0. This eigenspace is spanned by $b_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $b_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$. We are looking for an orthonormal basis (q_1, q_2) of this subspace. The Gram-Schmidt algorithm (page 234, but we are using different notations) proceeds by setting $\widetilde{b_1} = b_1$ and $\widetilde{b_2} = b_2 - \frac{\widetilde{b_1}^T b_2}{\widetilde{b_1}^T b_1} b_1$, and then setting $q_1 = \frac{\widetilde{b_1}}{\left|\left|\widetilde{b_1}\right|\right|}$ and $q_2 = \frac{\widetilde{b_2}}{\left|\left|\widetilde{b_2}\right|\right|}$. So

$\widetilde{b_1} = b_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and

$$\widetilde{b_2} = b_2 - \frac{\widetilde{b_1}^T b_2}{\widetilde{b_1}^T b_1} b_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \frac{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}.$$

Finally, $q_1 = \dfrac{\widetilde{b_1}}{\left|\left|\widetilde{b_1}\right|\right|} = \dfrac{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}}{\sqrt{2}} = \begin{pmatrix} -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \\ 0 \end{pmatrix}$ and $q_2 = \dfrac{\widetilde{b_2}}{\left|\left|\widetilde{b_2}\right|\right|} = \dfrac{\begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}}{\frac{1}{2}\sqrt{6}} =$ $\begin{pmatrix} -\frac{1}{6}\sqrt{6} \\ -\frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{6} \end{pmatrix}$.]

ad problem 3: We need to show that any two SVDs of A with the same Σ are obtained from each other in such a way.

So let $A = U\Sigma V^T$ and $A = \widetilde{U}\Sigma\widetilde{V}^T$ be two SVDs of A with the same Σ. Thus, U, V, $\widetilde{U}$ and $\widetilde{V}$ are orthogonal, and Σ is diagonal.

The matrix A is $n \times n$. Thus, the matrices U, V, $\widetilde{U}$, $\widetilde{V}$ and Σ are all $n \times n$. Let $\eta_1, \eta_2, \ldots, \eta_n$ be the diagonal entries of Σ (in the order in which they appear in Σ). Notice that these entries $\eta_1, \eta_2, \ldots, \eta_n$ are the singular values $\sigma_1, \sigma_2, \ldots, \sigma_n$ of A, but not necessarily in the same order. We have $\Sigma = \operatorname{diag}(\eta_1, \eta_2, \ldots, \eta_n)$, so that $\Sigma^T = \Sigma$ and $\Sigma^T\Sigma = \Sigma^2 = \operatorname{diag}\left(\eta_1^2, \eta_2^2, \ldots, \eta_n^2\right)$.

The singular values $\sigma_1, \sigma_2, \ldots, \sigma_n$ of A are pairwise distinct (since $\sigma_1 > \sigma_2 > \cdots > \sigma_n$). Thus, the matrix A has n pairwise distinct singular values. Hence, the matrix $A^T A$ has n pairwise distinct nonzero eigenvalues (since the singular values of A are the square roots of the nonzero eigenvalues of $A^T A$). As a consequence, for each of these eigenvalues, the matrix $A^T A$ has exactly one (up to scaling) eigenvector (because if it had a more-than-1-dimensional eigenspace for some eigenvector, then the sum of the dimensions of the n eigenspaces would total to something greater than n, but this cannot happen). Hence, for every $i \in \{1, 2, \ldots, n\}$,

$$\text{the matrix } A^T A \text{ has exactly one (up to scaling) eigenvector for the eigenvalue } \eta_i^2. \tag{1}$$

Let $u_1, u_2, \ldots, u_n$ be the columns of U, let $v_1, v_2, \ldots, v_n$ be the columns of V, let $\widetilde{u}_1, \widetilde{u}_2, \ldots, \widetilde{u}_n$ be the columns of $\widetilde{U}$, and let $\widetilde{v}_1, \widetilde{v}_2, \ldots, \widetilde{v}_n$ be the columns of $\widetilde{V}$.

The algorithm for finding an SVD suggests that $v_1, v_2, \ldots, v_n$ are orthonormal eigenvectors of $A^T A$ corresponding to the eigenvalues $\eta_1^2, \eta_2^2, \ldots, \eta_n^2$. This so far is only a conjecture, because the matrix A (usually) has many SVDs, and we do not know if any of our two SVDs is the one constructed by the algorithm! But maybe this conjecture holds for every SVD?

It turns out that it does. The vectors $v_1, v_2, \ldots, v_n$ are indeed orthonormal eigenvectors of $A^T A$ corresponding to the eigenvalues $\eta_1^2, \eta_2^2, \ldots, \eta_n^2$. To see this, we only need to show that they are eigenvectors for these eigenvalues (because we already know that they are orthonormal[2]). So, we must show that for every $i \in \{1, 2, \ldots, n\}$, the vector v_i is an eigenvector of $A^T A$ corresponding to the eigenvalue η_i^2. To prove this, we recall that $A = U\Sigma V^T$, so that $A^T = V\Sigma^T U^T$ (by problem 1 **(a)**), so that

$$\underbrace{A^T}_{=V\Sigma^T U^T} \underbrace{A}_{=U\Sigma V^T} V = V\Sigma^T \underbrace{U^T U}_{\substack{=I \\ \text{(since } U \\ \text{is orthogonal)}}} \Sigma \underbrace{V^T V}_{\substack{=I \\ \text{(since } V \\ \text{is orthogonal)}}} = V\Sigma^T I \Sigma I = V \underbrace{\Sigma^T \Sigma}_{=\operatorname{diag}\left(\eta_1^2, \eta_2^2, \ldots, \eta_n^2\right)}$$
$$= V \operatorname{diag}\left(\eta_1^2, \eta_2^2, \ldots, \eta_n^2\right).$$

But v_i is the i-th column of V, so that

$$A^T A v_i = A^T A \left(\text{the } i\text{-th column of } V\right) = \left(\text{the } i\text{-th column of } \underbrace{A^T A V}_{=V \operatorname{diag}\left(\eta_1^2, \eta_2^2, \ldots, \eta_n^2\right)}\right)$$
$$= \left(\text{the } i\text{-th column of } V \operatorname{diag}\left(\eta_1^2, \eta_2^2, \ldots, \eta_n^2\right)\right)$$
$$= \eta_i^2 \underbrace{\left(\text{the } i\text{-th column of } V\right)}_{=v_i} = \eta_i^2 v_i.$$

So, yes, v_i is an eigenvector of $A^T A$ corresponding to the eigenvalue η_i^2. But the same argument (with $\widetilde{U}$ and $\widetilde{V}$ taking the roles of U and V) shows that $\widetilde{v}_i$, too, is an eigenvector of $A^T A$ corresponding to the eigenvalue η_i^2. So both v_i and $\widetilde{v}_i$ are eigenvectors of $A^T A$ corresponding to the eigenvalue η_i^2. But (1) shows that there exists only one such eigenvector, up to scaling. Hence, v_i and $\widetilde{v}_i$ are equal up to scaling. Since v_i and $\widetilde{v}_i$ are length-1 vectors (being columns of the orthogonal matrices V and $\widetilde{V}$), this leaves only two possibilities: either $\widetilde{v}_i = v_i$ or $\widetilde{v}_i = -v_i$. In other words, we have $\widetilde{v}_i = s_i v_i$ for some $s_i \in \{1, -1\}$.

We have shown this for every $i \in \{1, 2, \ldots, n\}$. In other words, for every $i \in \{1, 2, \ldots, n\}$, there exists an $s_i \in \{1, -1\}$ such that $\widetilde{v}_i = s_i v_i$. This shows that the matrix $\widetilde{V}$ is obtained from the matrix V by multiplying some columns by -1.

Now, we need to prove that the matrix $\widetilde{U}$ is obtained from the matrix U by multiplying the **same** columns by -1. In other words, we need to show that every $i \in \{1, 2, \ldots, n\}$ satisfies $\widetilde{u}_i = s_i v_i$ for the very same s_i that we have just constructed.

Here we need to use the assumption that all σ_i are nonzero (this would not be true otherwise, as I have painfully learned myself). This assumption shows that all the η_i are nonzero (remember that the $\eta_1, \eta_2, \ldots, \eta_n$ are the $\sigma_1, \sigma_2, \ldots, \sigma_n$ in a possibly different order). Now, fix $i \in \{1, 2, \ldots, n\}$. Then,

$$\left(\text{the } i\text{-th column of } AV\right) = A \underbrace{\left(\text{the } i\text{-th column of } V\right)}_{=v_i} = Av_i,$$

[2]because they are the columns of the orthogonal matrix V

so that

$$\begin{aligned}
Av_i &= \left(\text{the } i\text{-th column of } \underbrace{A}_{=U\Sigma V^T} V\right) = \left(\text{the } i\text{-th column of } U\Sigma \underbrace{V^T V}_{\substack{=I \\ \text{(since } V \\ \text{is orthogonal)}}}\right) \\
&= (\text{the } i\text{-th column of } U\Sigma I) = \left(\text{the } i\text{-th column of } U \underbrace{\Sigma}_{=\operatorname{diag}(\eta_1,\eta_2,\ldots,\eta_n)}\right) \\
&= (\text{the } i\text{-th column of } U \operatorname{diag}(\eta_1,\eta_2,\ldots,\eta_n)) \\
&= \eta_i \underbrace{(\text{the } i\text{-th column of } U)}_{=u_i} = \eta_i u_i,
\end{aligned}$$

and thus $u_i = \dfrac{1}{\eta_i} A v_i$. (We were allowed to divide by η_i since η_i is nonzero.) Similarly, $\widetilde{u}_i = \dfrac{1}{\eta_i} A \widetilde{v}_i$. Since $\widetilde{v}_i = s_i v_i$, this becomes $\widetilde{u}_i = \dfrac{1}{\eta_i} A s_i v_i = s_i \underbrace{\dfrac{1}{\eta_i} A v_i}_{=u_i} = s_i u_i$. This proves what we wanted to.

ad problem 4:

Let us call the matrix A. This matrix is symmetric, so it makes sense to speak of its positive definiteness.

Let us recall that a symmetric $n \times n$-matrix B is positive definite if and only if all n upper-left determinants of B are positive.[3] Applying this to our 3×3-matrix A, we see that A is positive definite if and only if its upper-left determinants

$$\det(1) = 1,$$
$$\det\begin{pmatrix} 1 & u \\ u & 1 \end{pmatrix} = 1 - u^2, \qquad \text{and}$$
$$\det\begin{pmatrix} 1 & u & 0 \\ u & 1 & u \\ 0 & u & 1 \end{pmatrix} = 1 - 2u^2$$

are positive. This boils down to the inequalities $1 - u^2 > 0$ and $1 - 2u^2 > 0$.

[3] Recall that the k-th *upper-left determinant* of B (for $1 \leq k \leq n$) is the determinant of the matrix obtained by taking only the first k rows and the first k columns of B. For instance, the 2-nd upper-left determinant of $\begin{pmatrix} 1 & 2 & 4 \\ 2 & 6 & 4 \\ 4 & 4 & 3 \end{pmatrix}$ is $\det\begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix} = 2$. The n-th upper-left determinant of B is $\det B$, and the 1-st upper-left determinant of B is the first entry of the first row of B.

These two inequalities hold if and only if we have $-\dfrac{\sqrt{2}}{2} < u < \dfrac{\sqrt{2}}{2}$. Hence, the answer is "for those u which satisfy $-\dfrac{\sqrt{2}}{2} < u < \dfrac{\sqrt{2}}{2}$".

ad problem 5:

Assume that the two symmetric matrices A and B are similar.

We know that any symmetric matrix can be diagonalized orthogonally. Thus, A can be diagonalized orthogonally; i.e., there exists an orthogonal matrix P such that $P^{-1}AP$ is diagonal. By suitably permuting the columns of P (which are an orthonormal set of eigenvectors of A), we can make sure that the diagonal entries of $P^{-1}AP$ appear **in decreasing order** (because once we have found n orthonormal eigenvectors for A, we can reorder them to match any order of the eigenvalues that we choose). So we conclude that there exists an orthogonal matrix P such that the matrix $P^{-1}AP$ is diagonal and its diagonal entries appear in decreasing order. Similarly, there exists an orthogonal matrix Q such that the matrix $Q^{-1}BQ$ is diagonal and its diagonal entries appear in decreasing order. Let us consider these P and Q.

The matrix $P^{-1}AP$ is diagonal, and thus its diagonal entries are its eigenvalues. But its eigenvalues are precisely the eigenvalues of A (since $P^{-1}AP$ is similar to A), and those in turn are precisely the eigenvalues of B (since A is similar to B), and those again are the eigenvalues of $Q^{-1}BQ$ (since B is similar ot $Q^{-1}BQ$), and these in turn are the diagonal entries of $Q^{-1}BQ$ (since $Q^{-1}BQ$ is a diagonal matrix). So the diagonal entries of $P^{-1}AP$ are the same as the diagonal entries of $Q^{-1}BQ$, possibly up to order. But their order is also the same (recall that the diagonal entries appear in decreasing order in both $P^{-1}AP$ and $Q^{-1}BQ$). Hence, the diagonal entries of $P^{-1}AP$ are the same as the diagonal entries of $Q^{-1}BQ$, entry by entry. Thus, $P^{-1}AP = Q^{-1}BQ$ (since all other entries of these matrices are 0). Multiplying this equality by P from the left and by P^{-1} from the right, we obtain $A = PQ^{-1}BQP^{-1}$.

Now, take $M = PQ^{-1}$. Then,

$$M^TM = \underbrace{\left(PQ^{-1}\right)^T}_{=\left(Q^{-1}\right)^T P^T} PQ^{-1} = \underbrace{\left(Q^{-1}\right)^T}_{=\left(Q^T\right)^{-1}} \underbrace{P^TP}_{\substack{=I \\ \text{(since } P \text{ is} \\ \text{orthogonal)}}} Q^{-1} = \left(Q^T\right)^{-1} IQ^{-1}$$

$$= \left(Q^T\right)^{-1} Q^{-1} = \left(\underbrace{QQ^T}_{\substack{=I \\ \text{(since } Q \text{ is} \\ \text{orthogonal)}}} \right)^{-1} = I^{-1} = I,$$

so that M is orthogonal. Also, $A = \underbrace{PQ^{-1}}_{=M} B \underbrace{QP^{-1}}_{=\left(PQ^{-1}\right)^{-1}=M^{-1}} = MBM^{-1}$. So we are done.

ad problem 6:

(a) Assume that A is invertible. (If B is invertible, you can just switch A with B and do the same argument.) Then, $A(BA)A^{-1} = AB\underbrace{\left(AA^{-1}\right)}_{=I} = ABI = AB$.

Thus, BA is similar to AB, and we are done.

(b) No. For an example, take $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Then, $AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. The matrix BA is the zero matrix, which is only similar to itself, and thus not similar to the nonzero matrix AB.

ad problem 7:

I will give very little detail on this problem. See the solution to Exercise 0.2 on Darij's review problems for exam #3[4] for an example of how to argue linearity.

(a) The map T_1 is not linear, as it maps the zero polynomial 0 to the nonzero polynomial 2.

(b) The map T_2 is linear. (For example, it satisfies the second axiom because $T_2(f+g) = 2(f+g)(x) = \underbrace{2f(x)}_{=T_2(f)} + \underbrace{2g(x)}_{=T_2(g)} = T_2(f) + T_2(g)$.)

(c) The map T_3 is linear.

(d) The map T_4 is linear.

(e) The map T_5 is linear.

(f) The map T_6 is not linear. (It fails the second axiom, because $T_6(1+1) \neq T_6(1) + T_6(1)$. Indeed, $T_6(1+1) = T_6(2) = 2^2 = 4$ whereas $T_6(1) = 1^2 = 1$.)

(g) The map T_7 is linear. (This is mainly due to the fact that $(f+g)'' = f'' + g''$ and $(\alpha f)'' = \alpha f''$ for a **scalar** α.)

(h) The map T_8 is linear.

[If you find some of these answer paradoxic, remind yourself that f (and not x) is what is being transformed by the maps! So linear means "linear in f", not "linear in x" (whatever this would be). Does adding two f's and then applying the map give the same result as applying the maps to the two f's and then adding the results? Does scaling an f and then applying the map give the same result as applying the map to f and then scaling the result? These are the questions you should be asking yourself here. The map T_5, for example, transforms a polynomial f by replacing every x in f by x^2. This is perfectly linear in f, although it produces polynomials with lots of squares inside.]

ad problem 8:

[4] http://web.mit.edu/18.06/www/Fall14/Midterm3ReviewF14_Darij.pdf

For this problem and also for the next two, all that is needed is in Definition 0.5 in Michael's recitation #10[5] and on the first two pages of Darij's review problems for exam #3[6]. I will thus be rather brief.

We have

$$\begin{aligned} T(1) &= (2+x)\,1 + \int_0^x 1dt = 2x+2 = 2\cdot 1 + 2\cdot x + 0\cdot x^2 + 0\cdot x^3; \\ T(x) &= (2+x)\,x + \int_0^x t dt = \frac{3}{2}x^2 + 2x = 0\cdot 1 + 2\cdot x + \frac{3}{2}\cdot x^2 + 0\cdot x^3; \\ T\left(x^2\right) &= (2+x)\,x^2 + \int_0^x t^2 dt = \frac{4}{3}x^3 + 2x^2 = 0\cdot 1 + 0\cdot x + 2\cdot x^2 + \frac{4}{3}\cdot x^3. \end{aligned}$$

Thus, the matrix representing the map T with respect to our basis is

$$\begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & \frac{3}{2} & 2 \\ 0 & 0 & \frac{4}{3} \end{pmatrix}.$$

ad problem 9:

Denote the four matrices which form this basis by v_1, v_2, v_3, v_4 (so $v_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $v_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $v_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$). Note that **these matrices now serve as vectors**. You have worked with column vectors and row vectors all the time, but the general meaning of the word "vector" is "element in a vector space", and such elements can be numbers, column-vectors, polynomials or (as you are seeing here) matrices.

The map T takes a matrix and transposes it. So

$$\begin{aligned} T(v_1) &= T\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}^T = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = 1v_1 + 0v_2 + 0v_3 + 0v_4; \\ T(v_2) &= T\left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = 0v_1 + 0v_2 + 1v_3 + 0v_4; \\ T(v_3) &= T\left(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}^T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = 0v_1 + 1v_2 + 0v_3 + 0v_4; \\ T(v_4) &= T\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}^T = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = 0v_1 + 0v_2 + 0v_3 + 1v_4. \end{aligned}$$

[5] http://web.mit.edu/18.06/www/Fall14/Recitation10_Michael.pdf

[6] http://web.mit.edu/18.06/www/Fall14/Midterm3ReviewF14_Darij.pdf

Thus, the matrix representing the map T with respect to our basis is the 4×4-matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

ad problem 10:

One possible choice for a basis is (v_1, v_2), where $v_1 = (1, -1, 0)$ and $v_2 = (1, 0, -1)$. Then,

$$T(v_1) = T((1, -1, 0)) = (-1, 0, 1) = 0v_1 + (-1)v_2;$$
$$T(v_2) = T((1, 0, -1)) = (0, -1, 1) = 1v_1 + (-1)v_2.$$

Thus, the matrix representing the map T with respect to this basis is

$$\begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

[You might have chosen a different basis, and obtained a different matrix. However, notice that your matrix must be similar to my matrix. In particular, it has the trace -1 and the same determinant 1.]