Solution to 1:

(a) Suppose that $\mathbf{s}$ and $\mathbf{r}$ are vectors dependent on $\mathbf{x}$, i.e. $\mathbf{s} = \mathbf{s}(\mathbf{x})$ and $\mathbf{r} = \mathbf{r}(\mathbf{x})$. Then

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{s}^T\mathbf{r}) = \left(\frac{\partial \mathbf{s}}{\partial \mathbf{x}}\right)^T \mathbf{r} + \left(\frac{\partial \mathbf{r}}{\partial \mathbf{x}}\right)^T \mathbf{s}.$$

The above result can be verified component-wise. Applying this with $\mathbf{s} = \mathbf{x}/2$ and $\mathbf{r} = A\mathbf{x}$, we get:

$$\frac{\partial}{\partial \mathbf{x}}\left(\frac{1}{2}\mathbf{x}^T A\mathbf{x}\right) = \frac{1}{2}(A + A^T)\mathbf{x} = A\mathbf{x}.$$

The above needed the well-known result $(A\mathbf{x})^T = \mathbf{x}^T A^T$, which can be seen by components, and that $A = A^T$ (since A specified as symmetric).

For the second term, $\frac{\partial}{\partial \mathbf{x}}\mathbf{b}^T\mathbf{x} = \mathbf{b}$ can be seen by setting $\mathbf{s} = \mathbf{b}$ and $\mathbf{r} = \mathbf{x}$. So in conclusion, we have

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}.$$

(b) The result asked for in the question is known as the chain rule.

By definition,

$$\nabla g(f(\mathbf{x})) = \begin{pmatrix} \frac{\partial}{\partial x_1} g(f(\mathbf{x})) \\ \vdots \\ \frac{\partial}{\partial x_n} g(f(\mathbf{x})) \end{pmatrix}$$

When applying the partial derivative $\partial/\partial x_i$ to the function $g(f(\mathbf{x}))$, we can use the single variable chain rule because we are fixing all the other variables, so that the function $f(\mathbf{x})$ is treated as a single variable function.

Using the single variable chain rule, we have

$$\frac{\partial}{\partial x_i} g(f(\mathbf{x})) = g'(f(\mathbf{x}))\frac{\partial f}{\partial x_i}.$$

Applying this to all entries of $\nabla g(f(\mathbf{x}))$, we get:

$$\nabla g(f(\mathbf{x})) = \begin{pmatrix} g'(f(\mathbf{x}))\frac{\partial f}{\partial x_1} \\ \vdots \\ g'(f(\mathbf{x}))\frac{\partial f}{\partial x_n} \end{pmatrix} = g'(f(\mathbf{x}))\nabla f(\mathbf{x}).$$

(c) The Hessian operator can be expressed as the following:

$$\begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \left(\frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n}\right) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} & \frac{\partial^2}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} \\ \frac{\partial^2}{\partial x_2 \partial x_1} & \frac{\partial^2}{\partial x_2^2} & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} & \frac{\partial^2}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2}{\partial x_n^2} \end{pmatrix}.$$

So using our answer from part (a), we have:

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} (A\mathbf{x} + \mathbf{b})^T = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \mathbf{x}^T A^T + \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \mathbf{b}^T = \mathbf{I}A^T + \mathbf{0} = A^T = A$$

(d) Define the following function:

$$\begin{aligned} h : \mathbb{R}^n &\to \mathbb{R} \\ \mathbf{x} &\mapsto a^T\mathbf{x} \end{aligned}$$

This function is differentiable, since $h(\mathbf{x}+\epsilon) - h(\mathbf{x}) = a^T(\mathbf{x}+\epsilon) - a^T\mathbf{x} = a^T\mathbf{x} + a^T\epsilon - a^T\mathbf{x} = a^T\epsilon$, so the gradient of h is a. We can therefore apply the chain rule (from part (b)):

$$\triangledown f(\mathbf{x}) = \triangledown g(h(\mathbf{x})) = g'(h(\mathbf{x}))\triangledown h(\mathbf{x}) = g'(a^T\mathbf{x})a$$

To find the Hessian, we do as before and apply the gradient operator to the transpose of the gradient vector:

$$\begin{aligned}
\triangledown^2 f(\mathbf{x}) &= \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \begin{pmatrix} g'(a^T\mathbf{x})a_1 & g'(a^T\mathbf{x})a_2 & \cdots & g'(a^T\mathbf{x})a_n \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial}{\partial x_1} g'(a^T\mathbf{x})a_1 & \frac{\partial}{\partial x_1} g'(a^T\mathbf{x})a_2 & \cdots & \frac{\partial}{\partial x_1} g'(a^T\mathbf{x})a_n \\ \frac{\partial}{\partial x_2} g'(a^T\mathbf{x})a_1 & \frac{\partial}{\partial x_2} g'(a^T\mathbf{x})a_2 & \cdots & \frac{\partial}{\partial x_2} g'(a^T\mathbf{x})a_n \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n} g'(a^T\mathbf{x})a_1 & \frac{\partial}{\partial x_n} g'(a^T\mathbf{x})a_2 & \cdots & \frac{\partial}{\partial x_n} g'(a^T\mathbf{x})a_n \end{pmatrix} \\
&= \begin{pmatrix} g''(a^T\mathbf{x})a_1^2 & g''(a^T\mathbf{x})a_1a_2 & \cdots & g''(a^T\mathbf{x})a_1a_n \\ g''(a^T\mathbf{x})a_2a_1 & g''(a^T\mathbf{x})a_2^2 & \cdots & g''(a^T\mathbf{x})a_2a_n \\ \vdots & \vdots & \ddots & \vdots \\ g''(a^T\mathbf{x})a_na_1 & g''(a^T\mathbf{x})a_na_2 & \cdots & g''(a^T\mathbf{x})a_n^2 \end{pmatrix} \\
&= g''(a^T\mathbf{x})aa^T
\end{aligned}$$

Solution to 2:

(a) $A = \mathbf{z}\mathbf{z}^T$ is symmetric, since $A_{ij} = z_i z_j = z_j z_i = A_{ji}$.

Now pick an arbitrary vector $\mathbf{x} \in \mathbb{R}^n$. Need to show that $\mathbf{x}^T A \mathbf{x} \geq 0$.

$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= \mathbf{x}^T \mathbf{z}\mathbf{z}^T \mathbf{x} \\ &= (\mathbf{z}^T \mathbf{x})^T (\mathbf{z}^T \mathbf{x}) \\ &= (\mathbf{z}^T \mathbf{x})^2 \\ &>= 0 \end{aligned}$$

(b) If $A = \mathbf{z}\mathbf{z}^T$ where $z \in \mathbb{R}^n$ and non-zero, then the rank of A is 1, because each column is a multiple of each other: by definition of A, column i is $z_i \mathbf{z}$, and column j is $z_j \mathbf{z}$. Importantly $\mathbf{z} \neq \mathbf{0}$ means the rank isn't 0

The null-space of A is the set of all vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. Specifically, $\{(x_1 \quad x_2 \quad \ldots \quad x_n)^T \in \mathbb{R}^n \mid x_1 z_1 + x_2 z_2 + \cdots + x_n z_n = 0\}$. Since not all z_i are zero, this space has dimension $n - 1$.

(c) The answer is in the positive. Firstly, BAB^T is symmetric, since $(BAB^T)^T = (B^T)^T A^T B^T = BAB^T$. Here we used that A is symmetric.

Now pick an arbitrary vector $\mathbf{x} \in \mathbb{R}^m$. Need to show that $\mathbf{x}^T BAB^T \mathbf{x} \geq 0$.

$$\mathbf{x}^T BAB^T \mathbf{x} = (B^T \mathbf{x})^T AB^T \mathbf{x}$$

Note that $B^T \mathbf{x}$ is just a vector in $\mathbb{R}^n$, so let $\mathbf{z} = B^T \mathbf{x}$. By assumption, A is positive semi-definite, so $\mathbf{z}^T A \mathbf{z} \geq 0$. In conclusion

$$\begin{aligned} \mathbf{x}^T BAB^T \mathbf{x} &= (B^T \mathbf{x})^T AB^T \mathbf{x} \\ &= \mathbf{z}^T A \mathbf{z} \\ &\geq 0 \end{aligned}$$

so BAB^T is positive semi-definite.

Solution to 3:

(a) Suppose that $A = T\Lambda T^{-1}$ for invertible T and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

T is by definition the invertible matrix that maps e_i to $t^{(i)}$. Hence $T^{-1}t^{(i)} = e_i$. Then $\Lambda e_i = \lambda_i e_i$. Finally, $T(\lambda_i e_i) = \lambda_i T e_i = \lambda_i t^{(i)}$. In conclusion,

$$\begin{aligned} At^{(i)} &= T\Lambda T^{-1}t^{(i)} \\ &= T\Lambda e_i \\ &= T(\lambda_i e_i) \\ &= \lambda_i t^{(i)} \end{aligned}$$

(b) The idea is the same as in part (a). Only thing to note is that if U is orthogonal then it is invertible and $U^{-1} = U^T$. Then the argument should run exactly the same as in part (a).

(c) If A is positive semi-definite, then in particular A is symmetric. Hence we can use the Spectral Theorem and write down $A = U\Lambda U^T$ for some orthogonal matrix U and $\Lambda = \text{diag}(\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A))$.

From part b, we found $Au^{(i)} = \lambda_i(A)u^{(i)}$. Multiplying both sides by $(u^{(i)})^T$ we see that $(u^{(i)})^T Au^{(i)} = \lambda_i(A)(u^{(i)})^T u^{(i)}$. By assumption, $(u^{(i)})^T u^{(i)} = 1$ and $(u^{(i)})^T Au^{(i)} \geq 0$ for all i. Hence $\lambda_i(A) \geq 0$ for all i.