The point of this document is to derivate the gradient of the quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$. In particular, let $\mathbf{x}^T = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ be an n-dimensional column vector, and

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{(1)} \\ \mathbf{A}_{(2)} \\ \vdots \\ \mathbf{A}_{(n)} \end{pmatrix} = \left(\mathbf{A}^{(1)} \mid \mathbf{A}^{(2)} \mid \cdots \mid \mathbf{A}^{(n)} \right)$$

where $\mathbf{A}_{(i)}$ is the i^{th} row of $\mathbf{A}$ and $\mathbf{A}^{(i)}$ is the i^{th} column of $\mathbf{A}$.

The first thing we need to do is express $\mathbf{x}^T\mathbf{A}\mathbf{x}$ as an explicit function of the components of $\mathbf{x}$, $f(x_1, x_2, \ldots, x_n)$. To do this, we'll actually go through the matrix algebra and turn the product into a (double) sum. Observe,

$$\begin{aligned}
\mathbf{x}^T\mathbf{A}\mathbf{x} &= \mathbf{x}^T \begin{pmatrix} \mathbf{A}_{(1)} \\ \mathbf{A}_{(2)} \\ \vdots \\ \mathbf{A}_{(n)} \end{pmatrix} \mathbf{x} \\
&= \mathbf{x}^T \begin{pmatrix} \mathbf{A}_{(1)} \cdot \mathbf{x} \\ \mathbf{A}_{(2)} \cdot \mathbf{x} \\ \vdots \\ \mathbf{A}_{(n)} \cdot \mathbf{x} \end{pmatrix} \\
&= \mathbf{x}^T \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{pmatrix} \\
&= x_1 \sum_{j=1}^n a_{1j}x_j + x_2 \sum_{j=1}^n a_{2j}x_j + \cdots + x_n \sum_{j=1}^n a_{nj}x_j \\
&= \sum_{i=1}^n \left(x_i \sum_{j=1}^n a_{ij}x_j \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j \\
&\equiv f_{\mathbf{A}}(x_1, x_2, \ldots, x_n)
\end{aligned}$$

Now, recall that $\nabla f_{\mathbf{A}}(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$. So, let's look at one particular partial derivative to see if we can infer a pattern. Say, $n \geq 7$, let's consider $\partial f / \partial x_7$:

$$\begin{aligned}
\frac{\partial f}{\partial x_7} &= \frac{\partial}{\partial x_7} \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial}{\partial x_7} \left(x_i a_{ij} x_j \right)
\end{aligned}$$

and if we consider the four interesting cases for i and j:

$$
\begin{aligned}
&= \sum_{\substack{i\neq 7\\ j\neq 7}} \frac{\partial}{\partial x_7}\left(x_i a_{ij} x_j\right) + \sum_{\substack{i=7\\ j\neq 7}} \frac{\partial}{\partial x_7}\left(x_i a_{ij} x_j\right) + \sum_{\substack{i\neq 7\\ j=7}} \frac{\partial}{\partial x_7}\left(x_i a_{ij} x_j\right) + \sum_{\substack{i=7\\ j=7}} \frac{\partial}{\partial x_7}\left(x_i a_{ij} x_j\right) \\
&= \sum_{\substack{i\neq 7\\ j\neq 7}} 0 + \sum_{\substack{i=7\\ j\neq 7}} a_{7j} x_j + \sum_{\substack{i\neq 7\\ j=7}} a_{i7} x_i + \sum_{\substack{i=7\\ j=7}} 2a_{77} x_7 \\
&= \left(a_{77} x_7 + \sum_{\substack{i=7\\ j\neq 7}} a_{7j} x_j \right) + \left(a_{77} x_7 + \sum_{\substack{i\neq 7\\ j=7}} a_{i7} x_i \right) \\
&= \sum_{\substack{j=1\\ i\neq 7}}^{n} a_{7j} x_j + \sum_{\substack{i=1\\ j\neq 7}}^{n} a_{i7} x_i \\
&= \mathbf{x}^T \mathbf{A}_{(7)} + \mathbf{x}^T \mathbf{A}^{(7)} \\
&= \mathbf{x}^T \left(\mathbf{A}_{(7)} + \mathbf{A}^{(7)} \right)
\end{aligned}
$$

Hence,

$$\frac{\partial f}{\partial x_7} = \mathbf{x}^T \left(\mathbf{A}_{(7)} + \mathbf{A}^{(7)} \right)$$

However, since there was nothing particularly special about the seventh component, we may generalize this to conclude that

$$\frac{\partial f}{\partial x_k} = \mathbf{x}^T \left(\mathbf{A}_{(k)} + \mathbf{A}^{(k)} \right)$$

for $k = 1, 2, \ldots, n$. In particular,

$$
\begin{aligned}
\nabla f &= \left(\mathbf{x}^T \left(\mathbf{A}_{(1)} + \mathbf{A}^{(1)} \right), \mathbf{x}^T \left(\mathbf{A}_{(2)} + \mathbf{A}^{(2)} \right), \ldots, \mathbf{x}^T \left(\mathbf{A}_{(n)} + \mathbf{A}^{(n)} \right) \right) \\
&= \left(\mathbf{x}^T \mathbf{A}_{(1)}, \mathbf{x}^T \mathbf{A}_{(2)}, \ldots, \mathbf{x}^T \mathbf{A}_{(n)} \right) + \left(\mathbf{x}^T \mathbf{A}^{(1)}, \mathbf{x}^T \mathbf{A}^{(2)}, \ldots, \mathbf{x}^T \mathbf{A}^{(n)} \right) \\
&= \left(\mathbf{A}\mathbf{x} \right)^T + \left(\mathbf{A}^T \mathbf{x} \right)^T \\
&= \mathbf{x}^T \mathbf{A}^T + \mathbf{x}^T \mathbf{A} \\
&= \mathbf{x}^T \left(\mathbf{A}^T + \mathbf{A} \right)
\end{aligned}
$$

Thus, for a symmetric $\mathbf{A}$ (i.e. $\mathbf{A}$ such that $\mathbf{A}^T = \mathbf{A}$), we have that

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = 2\mathbf{x}^T \mathbf{A}$$