问题1：假设有函数f(y)，如果把x看成常数，y看成W的函数，如何根据函数对y的梯度值$\nabla_{y} f$计算函数对W的梯度值$\nabla_{W} f$？根据链式法则，由于w_{ij}只和y_j有关，和其他的y_k，$k \neq i$无关，因此有：

$$\frac{\partial f}{\partial w_{ij}}=\sum_{k=1}^{m}\frac{\partial f}{\partial y_k}\frac{\partial y_k}{\partial w_{ij}}=\sum_{k=1}^{m}\left(\frac{\partial f}{\partial y_k}\frac{\partial \sum_{l=1}^{n}\left(w_{kl}x_l\right)}{\partial w_{ij}}\right)=\frac{\partial f}{\partial y_i}\frac{\partial \sum_{l=1}^{n}\left(w_{il}x_l\right)}{\partial w_{ij}}=\frac{\partial f}{\partial y_i}x_j$$

对于W的所有元素有：

$$\begin{bmatrix}\frac{\partial f}{\partial w_{11}} & \cdots & \frac{\partial f}{\partial w_{1n}}\\ \cdots & \cdots & \cdots\\ \frac{\partial f}{\partial w_{m1}} & \cdots & \frac{\partial f}{\partial w_{mn}}\end{bmatrix}=\begin{bmatrix}\frac{\partial f}{\partial y_1}x_1 & \cdots & \frac{\partial f}{\partial y_1}x_n\\ \cdots & \cdots & \cdots\\ \frac{\partial f}{\partial y_m}x_1 & \cdots & \frac{\partial f}{\partial y_m}x_n\end{bmatrix}=\begin{bmatrix}\frac{\partial f}{\partial y_1}\\ \cdots\\ \frac{\partial f}{\partial y_m}\end{bmatrix}\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}$$

写成矩阵形式为：

$$\nabla_{W} f=\left(\nabla_{y} f\right)\mathrm{x}^{\mathrm{T}}$$

问题2：如果将W看成常数，y将看成x的函数，如何根据$\nabla_y f$计算$\nabla_x f$？由于任意的x_i和所有的y_i都有关系，根据链式法则有：

$$\frac{\partial f}{\partial x_i}=\sum_{j=1}^{m}\frac{\partial f}{\partial y_j}\frac{\partial y_j}{\partial x_i}=\sum_{j=1}^{m}\frac{\partial f}{\partial y_j}\frac{\partial\left(\sum_{k=1}^{n}w_{jk}x_k\right)}{\partial x_i}=\sum_{j=1}^{m}\frac{\partial f}{\partial y_i}w_{ji}=\begin{bmatrix}w_{1i} & \dots & w_{mi}\end{bmatrix}\nabla_y f$$

写成矩阵形式为：

$$\nabla_x f=\mathrm{W}^{\mathrm{T}}\nabla_y f$$

这是一个对称的结果，在计算函数映射时用矩阵W乘以向量x得到y，在求梯度时用矩阵W的转置乘以y的梯度得到x的梯度。

问题3：如果有向量到向量的映射：

$$\mathrm{y} = g(\mathrm{x})$$

写成分量形式为：

$$y_i = g(x_i)$$

在这里每个y_i只和对应的x_i有关，和其他所有$x_j, j \neq i$无关，且每个分量采用了相同的映射函数g。对于函数f(y)，如何根据$\nabla_{\mathrm{y}} f$计算$\nabla_{\mathrm{x}} f$？根据链式法则，由于每个y_i只和对应的x_i有关，有：

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial y_i} \frac{\partial y_i}{\partial x_i}$$

写成矩阵形式为：

$$\nabla_{\mathrm{x}} f = \nabla_{\mathrm{y}} f \odot g'(\mathrm{x})$$

即两个向量对应元素相乘，这种乘法在上一节已经介绍。

问题4：接下来我们考虑更复杂的情况，如果有下面的复合函数：

$$u = Wx$$
$$y = g(u)$$

其中g是向量对应元素一对一映射，即：

$$y_i = g(x_i)$$

如果有函数f(y)，如何根据$\nabla_y f$计算$\nabla_x f$？在这里有两层复合，首先是从x到u，然后是从u到y。根据问题2和问题3的结论，有：

$$\nabla_x f = W^T(\nabla_u f) = W^T\left((\nabla_y f)\odot g'(u)\right)$$

问题5：x是n维向量，y是m维向量，有映射y=g(x)，即：

$$y_i = g_i\left(x_1, x_2, \ldots, x_n\right), i = 1, \ldots, m$$

这里的映射方式和上面介绍的不同。对于向量y的每个分量y_i，映射函数g_i不同，而且y_i和向量x的每个分量x_j有关。对于函数f(y)，如何根据$\nabla_y f$计算$\nabla_x f$？根据链式法则，由于任何的y_i和任何的x_j都有关系，因此有：

$$\frac{\partial f}{\partial x_j} = \sum_{i=1}^{m} \frac{\partial f}{\partial y_i} \frac{\partial y_i}{\partial x_j} = \begin{bmatrix} \frac{\partial y_1}{\partial x_j} & \ldots & \frac{\partial y_m}{\partial x_j} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial y_1} \\ \ldots \\ \frac{\partial f}{\partial y_m} \end{bmatrix}$$

对于所有元素有：

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \ldots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \ldots & \frac{\partial y_m}{\partial x_1} \\ \ldots & \ldots & \ldots \\ \frac{\partial y_1}{\partial x_n} & \ldots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial y_1} \\ \ldots \\ \frac{\partial f}{\partial y_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \ldots & \frac{\partial y_1}{\partial x_n} \\ \ldots & \ldots & \ldots \\ \frac{\partial y_m}{\partial x_1} & \ldots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \frac{\partial f}{\partial y_1} \\ \ldots \\ \frac{\partial f}{\partial y_m} \end{bmatrix}$$

写成矩阵形式有：

$$\nabla_{\mathrm{x}} f = \left(\frac{\partial \mathrm{y}}{\partial \mathrm{x}}\right)^{\mathrm{T}} \nabla_{\mathrm{y}} f$$

其中$\frac{\partial \mathrm{y}}{\partial \mathrm{x}}$为雅可比矩阵。对于如下向量到向量的映射函数：

$$\mathrm{y} = f(\mathrm{x})$$

其中向量$\mathrm{x} \in \mathbb{R}^n$，向量$\mathrm{y} \in \mathbb{R}^m$这个映射写成分量形式为：

$$y_i = f_i(\mathrm{x})$$

即输出向量的每个分量是输入向量的函数。雅克比矩阵定义为输出向量的每个分量对输入向量的每个分量的偏导数构成的矩阵：

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \ldots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \ldots & \frac{\partial y_2}{\partial x_n} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \ldots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

这是一个m行n列的矩阵，每一行为一个多元函数的梯度。对于如下向量映射函数：

$$u = x^2 + 2xy + z$$
$$v = x - y^2 + z^2$$

它的雅克比矩阵为：

$$\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \end{bmatrix} = \begin{bmatrix} 2x+2y & 2x & 1 \\ 1 & -2y & 2z \end{bmatrix}$$

前面介绍的几个问题都是这个映射的特例。