# 矩阵、向量求导笔记

## 符号表示

- 标量用普通小写字母或希腊字母表示，如$t, \alpha$等；
- 用字母表示的向量默认为列向量，用粗体小写字母或粗体希腊字母表示，如$\mathbf{x}$等；行向量需写作列向量的转置，如$\mathbf{x^T}$等。其元素记作$x_i$（对于列向量）或$x_j$（对于行向量）。
- 矩阵用大写字母表示，如$\mathbf{A}$等，其元素记作$a_{ij}$。
- 有特殊说明的除外。

## 关于求导结果的约定

矩阵求导本身有很多争议，例如：

- 对于求导结果是否需要转置？
  - 不同教材对此处理的结果不一样。本文**以不转置为准**，即求导结果与原矩阵/向量同型。
- 矩阵对向量、向量对矩阵、矩阵对矩阵求导的结果是什么？
  - 这个问题非常混乱，不过主流的有两种处理方法：一是将矩阵进行平铺，但可能会遇到乘法法则不成立的问题。二是Magnus主张的方法，将矩阵抻成一个向量，然后再用向量求导的方法做。我个人觉得这些方法都不好，计算繁琐（比如要用Kronecker积构造等等）而且很多好的性质（比如链式法则）等都很难保存下来。
  - 事实上，如果把导数看成映射前空间到映射后空间的线性算子，那么应该本题目中求导的结果应当是三阶甚至四阶张量（我认为Fréchet导数的定义非常科学，可以参见维基百科 ），而张量运算本身又非常麻烦而且同样存在巨大的争议。
  - 考虑到机器学习中，事实上几乎不会遇到这几种情形的求导，我们不妨简单粗暴地认为**这三种情形下导数没有定义。**

综上所述，本文进行如下约定：

- 矩阵/向量值函数对实数的导数：
  - 要点：求导结果与函数值同型，且每个元素就是函数值的相应分量对自变量$x$求导
  - 若函数$\mathbf{F}:\mathbf{R}\to\mathbf{R^{m\times n}}$，则$\partial\mathbf{F}/\partial x$也是一个$m\times n$维矩阵，且$(\partial\mathbf{F}/\partial x)_{ij}=\partial f_{ij}/\partial x$。
  - 若函数$\mathbf{f}:\mathbf{R}\to\mathbf{R^{m\times 1}}$，则$\partial\mathbf{f}/\partial x$也是一个m维列向量，且$(\partial\mathbf{f}/\partial x)_i=\partial f_i/\partial x$。
  - 若函数$\mathbf{f^T}:\mathbf{R}\to\mathbf{R^{1\times n}}$，则$\partial\mathbf{f^T}/\partial x$也是一个n维行向量，且$(\partial\mathbf{f^T}/\partial x)_j=\partial f_j/\partial x$。
- 实值函数对矩阵/向量的导数：
  - 要点：求导结果与自变量同型，且每个元素就是$f$对自变量的相应分量求导
  - 若函数$f:\mathbf{R^{m\times n}}\to\mathbf{R}$，则$\partial f/\partial\mathbf{X}$也是一个$m\times n$维矩阵，且$(\partial f/\partial\mathbf{X})_{ij}=\partial f/\partial x_{ij}$。
  - 若函数$f:\mathbf{R^{m\times 1}}\to\mathbf{R}$，则$\partial f/\partial\mathbf{x}$也是一个$m$维列向量，且$(\partial f/\partial\mathbf{x})_i=\partial f/\partial x_i$。
  - 若函数$f:\mathbf{R^{1\times n}}\to\mathbf{R}$，则$\partial f/\partial\mathbf{x^T}$也是一个$n$维行向量，且$(\partial f/\partial\mathbf{x^T})_j=\partial f/\partial x_j$。
- 向量值函数对向量的导数：
  - 若函数$\mathbf{f}:\mathbf{R^{1\times n}}\to\mathbf{R^{m\times 1}}$，则$\partial\mathbf{f}/\partial\mathbf{x^T}$是一个$m\times n$维矩阵，且$(\partial\mathbf{f}/\partial\mathbf{x^T})_{ij}=\partial f_i/\partial x_j$。
  - 若函数$\mathbf{f}:\mathbf{R^{m\times 1}}\to\mathbf{R^{1\times n}}$，则$\partial\mathbf{f^T}/\partial\mathbf{x}$是一个$n\times m$维矩阵，定义为$\partial\mathbf{f^T}/\partial\mathbf{x}=(\partial\mathbf{f}/\partial\mathbf{x^T})^\mathbf{T}$。也即有$(\partial\mathbf{f^T}/\partial\mathbf{x})_{ij}=\partial f_j/\partial x_i$。
- 梯度、Hessian矩阵和劈形算子$\nabla$：
  - 有时也将矩阵/向量求导的结果用劈形算子表示，即：$\nabla_x f=\partial f/\partial x$（此式中$x$和$f$代表任意维度的向量或矩阵）。在求导的变量比较明确时，$\nabla$的下标可以省略，简记为$\nabla f$；
  - 对于一个实函数$f:\mathbf{R^{m\times 1}}\to\mathbf{R}$，其梯度规定为$m$维列向量$\nabla f=\mathbf{grad}f=\dfrac{\partial f}{\partial\mathbf{x}}$，Hessian矩阵规定为$\nabla^2 f=\nabla\nabla f=\dfrac{\partial(\nabla f)}{\partial\mathbf{x^T}}=\dfrac{\partial^2 f}{\partial\mathbf{x}\partial\mathbf{x^T}}$，是一个$m\times m$的矩阵。此时两个$\nabla$符号理解成分别对$\mathbf{x}$和$\mathbf{x^T}$求导（可交换顺序）。

- 对于一个实函数$f:\mathbf{R^{m\times n}}\to\mathbf{R}$，其梯度规定为$m\times n$维矩阵$\nabla f=\dfrac{\partial f}{\partial \mathbf{X}}$，Hessian矩阵不作定义。

# 对上述约定的理解

- 上面的定义始终满足转置关系：
  - 即：$\nabla_x f=(\nabla_{x^\mathbf{T}}f^\mathbf{T})^\mathbf{T}$（其中$x,f$代表任意维度的向量或矩阵）
  - 例如$\nabla_\mathbf{X} f=(\nabla_{\mathbf{X}^\mathbf{T}}f)^\mathbf{T}$等（$f$为实数时$f^T=f$）
- 函数增量的线性主部与自变量增量的关系：
  - 矩阵/向量值函数对实数的导数：
    - $\delta\mathbf{F}\approx\delta x(\nabla\mathbf{F})$（右边是实数和矩阵的数乘）
    - $\delta\mathbf{f}\approx\delta x(\nabla\mathbf{f})$（右边是实数和向量的数乘）
  - 实值函数对矩阵/向量的导数：
    - $\delta f\approx\sum_{i,j}(\nabla f)_{ij}(\delta\mathbf{X})_{ij}=tr((\nabla f)^T\delta\mathbf{X})$
      - 此式用到的技巧非常重要：两个同型矩阵对应元素相乘再求和时常用上面第二个等号转化为迹，从而简化运算。
      - 从另一个角度讲，这是矩阵导数的另一种定义。即：对于函数$f(X):\mathbf{R^{m\times n}}\to\mathbf{R}$，若存在矩阵$\mathbf{A}$，使得$||\delta\mathbf{X}||\to 0$时（$||\cdot||$为任意范数），成立$\delta f=tr(\mathbf{A^T}\delta\mathbf{X})+o(||\delta\mathbf{X}||)$，则定义$\nabla f=\mathbf{A}$。
      - **矩阵乘积的迹是一个线性算子，它在矩阵空间中的地位相当于内积在$n$维欧式空间中的地位！**
    - $\delta f\approx(\nabla f)^T\delta\mathbf{x}$（右边是向量内积）
      - 此式可看做前一个式子的退化情形。
  - 向量值函数对向量的导数：
    - $\delta\mathbf{f}\approx(\nabla_{\mathbf{x^T}}\mathbf{f})\delta\mathbf{x}$
      - 此式即为重积分换元时用于坐标变换的Jacobian矩阵。
    - $\delta\mathbf{f}\approx(\nabla_\mathbf{x}\mathbf{f^T})^\mathbf{T}\delta\mathbf{x}$
      - 与前式实质相同。

# 常用公式

- 实值/向量值函数对向量求导（未作特殊说明即为对$\mathbf{x}$求梯度）：
  - 行向量对列向量求导：
    - $\nabla\ \mathbf{x^T}=\mathbf{I},\nabla(\mathbf{Ax})^\mathbf{T}=\mathbf{A^T}$
  - 列向量对行向量求导：
    - $\nabla_{\mathbf{x^T}}\ \mathbf{x}=\mathbf{I},\nabla_{\mathbf{x^T}}(\mathbf{Ax})=\mathbf{A}$
  - 向量内积的求导法则（重要）：
    - $\nabla(\mathbf{u^T v})=\nabla(\mathbf{u^T})\cdot\mathbf{v}+\nabla(\mathbf{v^T})\cdot\mathbf{u}$
    - 特别地，有：
      - $\nabla||\mathbf{x}||_2^2=\nabla(\mathbf{x^T x})=2\mathbf{x},\nabla(\mathbf{w^T x})=\mathbf{w}$
      - $\nabla(\mathbf{x^T A x})=(\mathbf{A}+\mathbf{A^T})\mathbf{x}$
  - 向量数乘求导公式（较重要）：
    - $\nabla_{\mathbf{x^T}}(\alpha(\mathbf{x})\mathbf{f}(\mathbf{x}))=\mathbf{f}(\mathbf{x})\nabla_{\mathbf{x^T}}\alpha(\mathbf{x})+\alpha(\mathbf{x})\nabla_{\mathbf{x^T}}\mathbf{f}(\mathbf{x})$
- 矩阵迹求导（未作特殊说明即为对$\mathbf{X}$求梯度，下同）：
  - 迹的基本性质：
    - 线性性质：$tr(\sum_i c_i\mathbf{A}_i)=\sum_i c_i tr(\mathbf{A}_i)$
    - 转置不变性：$tr(\mathbf{A})=tr(\mathbf{A^T})$
    - 轮换不变性：$tr(\mathbf{ABCD})=tr(\mathbf{DABC})=\cdots$
  - 基本公式：
    - $\nabla tr(\mathbf{A^T X})=\mathbf{A}$（可以逐元素求导验证。事实上就是矩阵导数的第二种定义）
  - 迹方法的核心公式：
    - $\nabla tr(\mathbf{XAX^TB})=\mathbf{B^TXA^T}+\mathbf{BXA}$
    - 这个公式非常重要。在推导最小二乘解等问题上都会遇到。公式的名字是我瞎起的，我不知道它叫什么名字。
- 其他矩阵求导公式（大部分可由迹求导快速推出，不必强记。$\mathbf{u,v,A,B}$为与$\mathbf{X}$无关的常量）：
  - $\nabla\mathbf{u^T X v}=\mathbf{uv^T}$

- $\nabla \mathbf{u^T X^T X u} = 2\mathbf{X u u^T}$
- $\nabla \mathbf{(Xu-v)^T(Xu-v)} = 2\mathbf{(Xu-v)u^T}$
- $\nabla ||\mathbf{XA^T - B}||^2_{\mathbf{Fro}} = 2\mathbf{(XA^T-B)A}$
  - 特别地，$\nabla ||\mathbf{X}||^2_{\mathbf{Fro}} = \nabla \mathbf{(X^T X)} = 2\mathbf{X}$（根据逐元素求导易证）
- $\nabla_{\mathbf{X}} |\mathbf{X}| = |\mathbf{X}|(\mathbf{X^{-1}})^{\mathbf{T}}$（可用逐元素求导 + 伴随矩阵的性质推导）
- 链式法则：
  - 设$\mathbf{U} = f(\mathbf{X})$，则：
    - $\dfrac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k,l} \dfrac{g(\mathbf{U})}{\partial u_{kl}} \dfrac{\partial u_{kl}}{\partial x_{ij}}$，或简写为$\dfrac{\partial g(\mathbf{U})}{\partial x_{ij}} = tr((\dfrac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^{\mathbf{T}} \dfrac{\partial \mathbf{U}}{\partial x_{ij}})$
- 线性变换的导数（可以直接用导数定义证。可以简化很多公式的推导过程）：
  - 设有$f(\mathbf{Y}) : \mathbf{R^{m\times p}} \to \mathbf{R}$及线性映射 $\mathbf{X} \mapsto \mathbf{Y} = \mathbf{AX+B} : \mathbf{R^{n\times p}} \to \mathbf{R^{m\times p}}$，则：
    - $\nabla_{\mathbf{X}}\ f(\mathbf{AX+B}) = \mathbf{A^T}\nabla_{\mathbf{Y}} f$
    - 向量的线性变换是上式的退化情形，即：$\nabla_{\mathbf{x}}\ f(\mathbf{Ax+b}) = \mathbf{A^T}\nabla_{\mathbf{y}} f$
    - 向量的线性变换还可以求二阶导：$\nabla^2_{\mathbf{x}}\ f(\mathbf{Ax+b}) = \mathbf{A^T}(\nabla^2_{\mathbf{y}} f)\mathbf{A}$
- 矩阵/向量对实数求导：
  - $(|\mathbf{F}|)'_x = |\mathbf{F}| tr(\mathbf{F^{-1}F'}_x)$
  - $(\ln|\mathbf{F}|)'_x = tr(\mathbf{F^{-1}F'_x})$
  - $\mathbf{(F^{-1})'}_x = -\mathbf{F^{-1}F'}_x\mathbf{F^{-1}}$（可对恒等式$\mathbf{FF^{-1}} = \mathbf{I}$两边同时求导来证明）

此外应注意到以下两条规律：

- 多个矩阵相乘时，其增量的线性主部等于自变量的每一次出现引起的增量之和。因此，可单独计算自变量每一次出现引起的导数变化，再把这些结果加起来。
  - 例如：

$$\delta(\mathbf{C_1F_1C_2F_2C_3}) = \delta(\mathbf{C_1F_1C_2F_{2c}C_3}) + \delta(\mathbf{C_1F_{1c}C_2F_2C_3}) \\ + \text{higher order infinitesimal}$$

，其中$\mathbf{F_{ic}}$表示将$\mathbf{F_i}$视为常数，而非自变量的函数。
  - 可以据此推导矩阵迹方法的核心公式：
    - $\nabla tr(\mathbf{XAX^TB}) = \nabla tr(\mathbf{X_cAX^TB}) + \nabla tr(\mathbf{XAX_c^TB}) = \mathbf{BXA + B^TXA^T}$
- 实数在与一堆矩阵、向量作数乘时可以随意移动位置。且实数乘行向量时，向量数乘与矩阵乘法（$1\times 1$矩阵和$1\times m$矩阵相乘）是兼容的。

# 要点

- 遇到相同下标求和就联想到矩阵乘法的定义，即$c_{ij} = \sum_j a_{ij}b_{jk}$。特别地，一维下标求和联想到向量内积$\sum_i u_iv_i = \mathbf{u^Tv}$，二维下标求和联想到迹$\sum_{ij} a_{ij}b_{ij} = tr(\mathbf{AB^T})$（$\mathbf{A,B}$应为同型矩阵）。
- 如果在一个求和式中，待求和项为矩阵的乘积，不要想着展开，而要按照上面的思路，看成分块矩阵的相乘！
- 向量的模长（或实数的平方和）转化为内积运算：$\sum_i x_i^2 = \mathbf{x^Tx}$。矩阵的F范数转化为迹运算：$||\mathbf{A}||^2_{\mathbf{Fro}} = tr(\mathbf{AA^T})$。
- 多个矩阵相乘时，多用矩阵迹的求导公式转化、循环移动各项！实数也可看成$1\times 1$矩阵的迹！

# 算例

- 最小二乘解推导：
  - 方法一：展开括号，再使用几个常用公式化简即可：
    - 

$$\begin{aligned} \nabla_{\mathbf{x}}||\mathbf{Ax-b}||_2^2 = &\quad \nabla_{\mathbf{x}}\mathbf{(Ax-b)^T(Ax-b)} \\ = &\quad \nabla_{\mathbf{x}}(\mathbf{x^TA^TAx - b^TAx - x^TA^Tb + b^Tb}) \\ = &\quad \nabla_{\mathbf{x}}(\mathbf{x^TA^TAx}) - 2\nabla_{\mathbf{x}}(\mathbf{b^TAx}) + \nabla_{\mathbf{x}}(\mathbf{b^Tb}) \\ = &\quad 2\mathbf{A^TAx} - 2\mathbf{A^Tb} + \mathbf{0} \\ = &\quad 2\mathbf{A^T(Ax-b)} \end{aligned}$$

- 方法二：使用线性变换的求导公式：
  - $$\begin{aligned}\nabla_{\mathbf{x}}||\mathbf{Ax}-\mathbf{b}||_2^2 = & \quad \mathbf{A^T}\nabla_{\mathbf{Ax-b}}||\mathbf{Ax}-\mathbf{b}||_\mathbf{2}^\mathbf{2} \\ = & \quad \mathbf{A^T}(2(\mathbf{Ax}-\mathbf{b})) \\ = & \quad 2\mathbf{A^T}(\mathbf{Ax}-\mathbf{b})\end{aligned}$$

- F范数的求导公式推导：
  - 方法一：先转化为迹，再裂项，最后通过恰当的轮换，用迹方法的核心公式处理。
    - $$\begin{aligned}\nabla||\mathbf{XA^T}-\mathbf{B}||_{\mathbf{Fro}}^2 = & \quad \nabla tr((\mathbf{XA^T}-\mathbf{B})^\mathbf{T}(\mathbf{XA^T}-\mathbf{B})) \\ = & \quad \nabla tr(\mathbf{AX^TXA^T}-\mathbf{B^TXA^T}-\mathbf{AX^TB}+\mathbf{B^TB}) \\ = & \quad \nabla tr(\mathbf{AX^TXA^T})-2tr(\mathbf{AX^TB})+tr(\mathbf{B^TB}) \\ = & \quad 2tr(\mathbf{XA^TAX^TI})-2tr(\mathbf{X^TBA})+\mathbf{0} \\ = & \quad 2(\mathbf{I^TX}(\mathbf{A^TA})^\mathbf{T}+\mathbf{IX}(\mathbf{A^TA}))-2\mathbf{BA} \\ = & \quad 2\mathbf{XA^TA}-2\mathbf{BA} \\ = & \quad 2(\mathbf{XA^T}-\mathbf{B})\mathbf{A}\end{aligned}$$
  - 方法二：用线性变换的求导公式证。（注意矩阵转置不改变其F范数，并且实值函数对$\mathbf{X}$和$\mathbf{X^T}$的导数互为转置）
    - $$\begin{aligned}\nabla||\mathbf{XA^T}-\mathbf{B}||_{\mathbf{Fro}}^2 = & \quad \nabla||\mathbf{AX^T}-\mathbf{B^T}||_{\mathbf{Fro}}^2 \\ = & \quad (\nabla_{\mathbf{X^T}}||\mathbf{AX^T}-\mathbf{B^T}||_{\mathbf{Fro}}^2)^\mathbf{T} \\ = & \quad (\mathbf{A^T}(2(\mathbf{AX^T}-\mathbf{B^T})))^\mathbf{T} \\ = & \quad 2(\mathbf{XA^T}-\mathbf{B})\mathbf{A}\end{aligned}$$
  - 方法三：根据定义逐元素地算，然后合并成向量、再合并成矩阵。（太原始、低效，不推荐）
- PRML (3.33)求导：
  - 题目：
    - 求$f(\mathbf{W})=\ln p(\mathbf{T}|\mathbf{X},\mathbf{W},\beta)=\text{const}-\dfrac{\beta}{2}\sum_n||\mathbf{t_n}-\mathbf{W^T}\phi(\mathbf{x_n})||_2^2$关于$\mathbf{W}$的导数。
  - 方法一：用矩阵的F范数推导：
    - $$\begin{aligned}\nabla f = & \quad \nabla\left(-\frac{\beta}{2}\sum_n||\mathbf{t_n}-\mathbf{W^T}\phi(\mathbf{x_n})||_2^2\right) \\ = & \quad -\frac{\beta}{2}\nabla||\mathbf{T^T}-\mathbf{W^T\Phi^T}||_{\mathbf{Fro}} \\ = & \quad -\frac{\beta}{2}\nabla||\mathbf{T}-\mathbf{\Phi W}||_{\mathbf{Fro}} \\ = & \quad -\frac{\beta}{2}\nabla||\mathbf{\Phi W}-\mathbf{T}||_{\mathbf{Fro}} \\ = & \quad -\frac{\beta}{2}\mathbf{\Phi^T}(2(\mathbf{\Phi W}-\mathbf{T})) \\ = & \quad -\beta\mathbf{\Phi^T}(\mathbf{\Phi W}-\mathbf{T})\end{aligned}$$
    - 上述几步的依据分别是：
      - 将若干个列向量拼成一个矩阵，因此它们的二范数平方和就等于大矩阵的F范数。
      - 矩阵转置不改变其F范数。
      - 矩阵数乘(-1)不改变其F范数。
      - 线性变换的求导公式 + F范数的求导公式。
      - 实数在和矩阵作数乘时位置可以任意移动。

- 于是求得$\mathbf{W}$的最大似然解为$\mathbf{W_{ML}}=\mathbf{\Phi^{\dagger}T}=(\mathbf{\Phi^T\Phi})^{-1}\mathbf{\Phi^T}$。

- 方法二：将向量二范数用内积代替，然后逐项展开，最后利用分块矩阵相乘消掉求和号：
  - 

$$
\begin{aligned}
\nabla f = & \nabla\left(-\frac{\beta}{2}\sum_n ||\mathbf{t_n}-\mathbf{W^T}\phi(\mathbf{x_n})||_2^2\right) \\
= & -\frac{\beta}{2}\nabla\left(\sum_n(\mathbf{t_n}-\mathbf{W^T}\phi(\mathbf{x_n}))^\mathbf{T}(\mathbf{t_n}-\mathbf{W^T}\phi(\mathbf{x_n}))\right) \\
= & -\frac{\beta}{2}\sum_n\{\nabla(\mathbf{t_n^T t_n})-2\nabla(\phi(\mathbf{x_n})^\mathbf{T}\mathbf{W}\mathbf{t_n}) \\
& +\nabla(\phi(\mathbf{x_n})^\mathbf{T}\mathbf{W}\mathbf{W^T}\phi(\mathbf{x_n}))\} \\
= & -\frac{\beta}{2}\sum_n\{\mathbf{0}-2\phi(\mathbf{x_n})\mathbf{t_n^T}+\nabla(\mathbf{W}\mathbf{I}\mathbf{W^T}\phi(\mathbf{x_n})\phi(\mathbf{x_n})^\mathbf{T})\} \\
= & -\frac{\beta}{2}\sum_n\{-2\phi(\mathbf{x_n})\mathbf{t_n^T}+(\phi(\mathbf{x_n})\phi(\mathbf{x_n})^\mathbf{T})^\mathbf{T}\mathbf{W}\mathbf{I^T}+(\phi(\mathbf{x_n})\phi(\mathbf{x_n})^\mathbf{T})\mathbf{W}\mathbf{I}\} \\
= & -\frac{\beta}{2}\sum_n\{-2\phi(\mathbf{x_n})\mathbf{t_n^T}+2\phi(\mathbf{x_n})\phi(\mathbf{x_n})^\mathbf{T}\mathbf{W}\mathbf{I}\} \\
= & -\beta\sum_n\{-\phi(\mathbf{x_n})\mathbf{t_n^T}+\phi(\mathbf{x_n})\phi(\mathbf{x_n})^\mathbf{T}\mathbf{W}\} \\
= & -\beta\sum_n\phi(\mathbf{x_n})\{-\mathbf{t_n^T}+\phi(\mathbf{x_n})^\mathbf{T}\mathbf{W}\} \\
= & -\beta\mathbf{\Phi^T}(\mathbf{\Phi W}-\mathbf{T})
\end{aligned}
$$

  - 注意最后一步的思考过程：
    - 将对$n$求和视为两个分块矩阵的乘积：
      - 第一个矩阵是分块行向量，共$1\times N$个块，且第n个分量是$\phi(\mathbf{x_n})$。因此第一个矩阵是$(\phi(\mathbf{x_1}),\phi(\mathbf{x_2}),\cdots,\phi(\mathbf{x_N}))=\mathbf{\Phi^T}$
      - 第二个矩阵是分块列向量，共$N\times 1$个块，且第n个分量是$-\mathbf{t_n^T}+\phi(\mathbf{x_n})^\mathbf{T}\mathbf{W}$。因此，第二个矩阵是：

$$
\begin{aligned}
\begin{pmatrix}-\mathbf{t_1^T}+\phi(\mathbf{x_1})^\mathbf{T}\mathbf{W}\\-\mathbf{t_2^T}+\phi(\mathbf{x_2})^\mathbf{T}\mathbf{W}\\\vdots\\-\mathbf{t_N^T}+\phi(\mathbf{x_N})^\mathbf{T}\mathbf{W}\end{pmatrix} = & \begin{pmatrix}\phi(\mathbf{x_1})^\mathbf{T}\mathbf{W}\\\phi(\mathbf{x_2})^\mathbf{T}\mathbf{W}\\\vdots\\\phi(\mathbf{x_N})^\mathbf{T}\mathbf{W}\end{pmatrix}-\begin{pmatrix}\mathbf{t_1^T}\\\mathbf{t_2^T}\\\vdots\\\mathbf{t_N^T}\end{pmatrix} \\
= & \begin{pmatrix}\phi(\mathbf{x_1})^\mathbf{T}\\\phi(\mathbf{x_2})^\mathbf{T}\\\vdots\\\phi(\mathbf{x_N})^\mathbf{T}\end{pmatrix}\mathbf{W}-\begin{pmatrix}\mathbf{t_1^T}\\\mathbf{t_2^T}\\\vdots\\\mathbf{t_N^T}\end{pmatrix} \\
= & \mathbf{\Phi W}-\mathbf{T}
\end{aligned}
$$

        ，注意第二个等式的推导过程中，第一项能够拆开是因为它被看做两个分块矩阵的乘积，两个分块矩阵分别由$N\times 1$和$1\times 1$个块组成。
  - 这种方法虽然比较繁琐，但是更具有一般性。