11.9

$\|\nabla f(\boldsymbol{x}')-\nabla f(\boldsymbol{x})\|_2^2 \leqslant L\|\boldsymbol{x}'-\boldsymbol{x}\|_2^2 \quad (\forall \boldsymbol{x},\boldsymbol{x}')$, [解析]：*L-Lipschitz*条件定义为：对于函数$f(\boldsymbol{x})$,若其任意定义域中的**x1**,**x2**，都存在L>0，使得$|f(\boldsymbol{x}1)-f(\boldsymbol{x}2)| \leq L|\boldsymbol{x}1-\boldsymbol{x}2|$。通俗理解就是，对于函数上的每一对点，都存在一个实数L，使得它们连线斜率的绝对值不大于这个L，其中最小的L称为*Lipschitz*常数。 将公式变形可以更好的理解：$\frac{\|\nabla f(\boldsymbol{x}')-\nabla f(\boldsymbol{x})\|_2^2}{\|\boldsymbol{x}'-\boldsymbol{x}\|_2^2} \leqslant L \quad (\forall \boldsymbol{x},\boldsymbol{x}')$, 进一步，如果$\boldsymbol{x}' \to \boldsymbol{x}$，即$\lim_{\boldsymbol{x}' \to \boldsymbol{x}} \frac{\|\nabla f(\boldsymbol{x}')-\nabla f(\boldsymbol{x})\|_2^2}{\|\boldsymbol{x}'-\boldsymbol{x}\|_2^2}$ “ *Lipschitz*连续”很常见，知乎有一个问答(https://www.zhihu.com/question/51809602) 对*Lipschitz*连续的解释很形象：以陆地为例， 连续就是说这块地上没有特别陡的坡；其中最陡的地方有多陡呢？这就是所谓的*Lipschitz*常数。

11.10

$$\hat{f}(x) \simeq f(x_k)+\langle \nabla f(x_k), x-x_k\rangle + \frac{L}{2}\|x-x_k\|^2$$

[推导]：

$$\begin{aligned}
\hat{f}(x) &\simeq f(x_k)+\langle \nabla f(x_k), x-x_k\rangle + \frac{L}{2}\|x-x_k\|^2 \\
&= f(x_k)+\langle \nabla f(x_k), x-x_k\rangle + \langle \frac{L}{2}(x-x_k), x-x_k\rangle \\
&= f(x_k)+\langle \nabla f(x_k) + \frac{L}{2}(x-x_k), x-x_k\rangle \\
&= f(x_k)+\frac{L}{2}\langle \frac{2}{L}\nabla f(x_k) + (x-x_k), x-x_k\rangle \\
&= f(x_k)+\frac{L}{2}\langle x-x_k+\frac{1}{L}\nabla f(x_k)+\frac{1}{L}\nabla f(x_k), x-x_k+\frac{1}{L}\nabla f(x_k)-\frac{1}{L}\nabla f(x_k)\rangle \\
&= f(x_k)+\frac{L}{2}\left\|x-x_k+\frac{1}{L}\nabla f(x_k)\right\|_2^2 - \frac{1}{2L}\|\nabla f(x_k)\|_2^2 \\
&= \frac{L}{2}\left\|x-(x_k-\frac{1}{L}\nabla f(x_k))\right\|_2^2 + const \qquad (\text{因为} f(x_k) \text{和} \nabla f(x_k) \text{是常数})
\end{aligned}$$

11.13

$\boldsymbol{x_{k+1}} = \underset{\boldsymbol{x}}{argmin} \frac{L}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$ [推导]：假设目标函数为$g(\boldsymbol{x})$，则

$$\begin{aligned}
g(\boldsymbol{x}) &= \frac{L}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \\
&= \frac{L}{2}\sum_{i=1}^{d}\left\|x^i-z^i\right\|_2^2 + \lambda\sum_{i=1}^{d}\left\|x^i\right\|_1 \\
&= \sum_{i=1}^{d}(\frac{L}{2}(x^i-z^i)^2+\lambda\left|x^i\right|)
\end{aligned}$$

由上式可见， $g(\boldsymbol{x})$可以拆成 d个独立的函 数，求解式(11.13)可以分别求解d个独立的目标函数。 针对目标函数$g(x^i)=\frac{L}{2}(x^i-z^i)^2+\lambda\left|x^i\right|$，通过求导求解极值： $\frac{dg(x^i)}{dx^i}=L(x^i-z^i)+\lambda sgn(x^i)$ 其中 $sgn(x^i)=\begin{cases} 1, & x^i>0 \\ -1, & x^i<0 \end{cases}$ 令导数为0，可得：$x^i=z^i-\frac{\lambda}{L}sgn(x^i)$可分为三种情况：

1. 当$z^i>\frac{\lambda}{L}$时：（1）假设此时的根$x^i<0$，则$sgn(x^i)=-1$，所以$x^i=z^i+\frac{\lambda}{L}>0$，与假设矛盾。（2）假设此时的根$x^i>0$，则$sgn(x^i)=1$，所以$x^i=z^i-\frac{\lambda}{L}>0$，成立。
2. 当$z^i<-\frac{\lambda}{L}$时：（1）假设此时的根$x^i>0$，则$sgn(x^i)=1$，所以$x^i=z^i-\frac{\lambda}{L}<0$，与假设矛盾。（2）假设此时的根$x^i<0$，则$sgn(x^i)=-1$，所以$x^i=z^i+\frac{\lambda}{L}<0$，成立。

3. 当$|z^i| < \frac{\lambda}{L}$时：（1）假设此时的根$x^i > 0$，则$sgn(x^i) = 1$，所以$x^i = z^i - \frac{\lambda}{L} < 0$，与假设矛盾。（2）假设此时的根$x^i < 0$，则$sgn(x^i) = -1$，所以$x^i = z^i + \frac{\lambda}{L} > 0$，与假设矛盾，此时$x^i = 0$为函数的极小值。 综上所述可得函数闭式解如下：$x_{k+1}^i = \begin{cases} z^i - \frac{\lambda}{L}, & \frac{\lambda}{L} < z^i \\ 0, & |z^i| \leqslant \frac{\lambda}{L} \\ z^i + \frac{\lambda}{L}, & z^i < -\frac{\lambda}{L} \end{cases}$

11.18

$$\begin{aligned}
\underset{\boldsymbol{B}}{min}\|\boldsymbol{X}-\boldsymbol{BA}\|_F^2 &= \underset{b_i}{min}\left\|\boldsymbol{X}-\sum_{j=1}^{k} b_j\alpha^j\right\|_F^2 \\
&= \underset{b_i}{min}\left\|\left(\boldsymbol{X}-\sum_{j\neq i} b_j\alpha^j\right)-b_i\alpha^i\right\|_F^2 \\
&= \underset{b_i}{min}\left\|\boldsymbol{E_i}-b_i\alpha^i\right\|_F^2
\end{aligned}$$

[推导]：此处只推导一下$BA = \sum_{j=1}^{k} \boldsymbol{b_j}\boldsymbol{\alpha^j}$，其中$\boldsymbol{b_j}$表示**B**的第j列，$\boldsymbol{\alpha^j}$表示**A**的第j行。 然后，用$b_j^i$，$\alpha_j^i$分别表示**B**和**A**的第i行第j列的元素，首先计算**BA**:

$$\begin{aligned}
\boldsymbol{BA} &= \begin{bmatrix} b_1^1 & b_2^1 & \cdot & \cdot & \cdot & b_k^1 \\ b_1^2 & b_2^2 & \cdot & \cdot & \cdot & b_k^2 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ b_1^d & b_2^d & \cdot & \cdot & \cdot & b_k^d \end{bmatrix}_{d\times k} \cdot \begin{bmatrix} \alpha_1^1 & \alpha_2^1 & \cdot & \cdot & \cdot & \alpha_m^1 \\ \alpha_1^2 & \alpha_2^2 & \cdot & \cdot & \cdot & \alpha_m^2 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ \alpha_1^k & \alpha_2^k & \cdot & \cdot & \cdot & \alpha_m^k \end{bmatrix}_{k\times m} \\
&= \begin{bmatrix} \sum_{j=1}^k b_j^1\alpha_1^j & \sum_{j=1}^k b_j^1\alpha_2^j & \cdot & \cdot & \cdot & \sum_{j=1}^k b_j^1\alpha_m^j \\ \sum_{j=1}^k b_j^2\alpha_1^j & \sum_{j=1}^k b_j^2\alpha_2^j & \cdot & \cdot & \cdot & \sum_{j=1}^k b_j^2\alpha_m^j \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ \sum_{j=1}^k b_j^d\alpha_1^j & \sum_{j=1}^k b_j^d\alpha_2^j & \cdot & \cdot & \cdot & \sum_{j=1}^k b_j^d\alpha_m^j \end{bmatrix}_{d\times m}
\end{aligned}$$

然后计算$\boldsymbol{b_j}\boldsymbol{\alpha^j}$:

$$
\boldsymbol{b_j \alpha^j} = \begin{bmatrix} b_1^j \\ b_w^j \\ \cdot \\ \cdot \\ \cdot \\ b_d^j \end{bmatrix} \cdot \begin{bmatrix} \alpha_1^j & \alpha_2^j & \cdot & \cdot & \cdot & \alpha_m^j \end{bmatrix}
$$

$$
= \begin{bmatrix} b_j^1 \alpha_1^j & b_j^1 \alpha_2^j & \cdot & \cdot & \cdot & b_j^1 \alpha_m^j \\ b_j^2 \alpha_1^j & b_j^2 \alpha_2^j & \cdot & \cdot & \cdot & b_j^2 \alpha_m^j \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ b_j^d \alpha_1^j & b_j^d \alpha_2^j & \cdot & \cdot & \cdot & b_j^d \alpha_m^j \end{bmatrix}_{d \times m}
$$

求和可得：

$$
\sum_{j=1}^{k} \boldsymbol{b_j \alpha^j} = \sum_{j=1}^{k} \left(\begin{bmatrix} b_1^j \\ b_w^j \\ \cdot \\ \cdot \\ \cdot \\ b_d^j \end{bmatrix} \cdot \begin{bmatrix} \alpha_1^j & \alpha_2^j & \cdot & \cdot & \cdot & \alpha_m^j \end{bmatrix} \right)
$$

$$
= \begin{bmatrix} \sum_{j=1}^{k} b_j^1 \alpha_1^j & \sum_{j=1}^{k} b_j^1 \alpha_2^j & \cdot & \cdot & \cdot & \sum_{j=1}^{k} b_j^1 \alpha_m^j \\ \sum_{j=1}^{k} b_j^2 \alpha_1^j & \sum_{j=1}^{k} b_j^2 \alpha_2^j & \cdot & \cdot & \cdot & \sum_{j=1}^{k} b_j^2 \alpha_m^j \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ \sum_{j=1}^{k} b_j^d \alpha_1^j & \sum_{j=1}^{k} b_j^d \alpha_2^j & \cdot & \cdot & \cdot & \sum_{j=1}^{k} b_j^d \alpha_m^j \end{bmatrix}_{d \times m}
$$