General definitions

1. $\sigma = 0.1$
2. $h : \mathbb{R}^2 \to \mathbb{R} : x \mapsto h(x) = \exp\left(\frac{-\|x\|^2}{\sigma^2}\right)$
3. $\varphi : \mathbb{R}^2 \to \mathbb{R}^{n^2} : x \mapsto \varphi(x) = \begin{bmatrix} h(p_1 - x) \\ \vdots \\ h(p_{n^2} - x) \end{bmatrix}$
4. $\Phi : \mathbb{R}^{2\times K} \to \mathbb{R}^{n^2} : X \mapsto \Phi(X) = \sum_{k=1}^{K} \varphi(x_k)$
5. $f : \mathbb{R}^{2\times K} \to \mathbb{R} : X \mapsto f(X) = \frac{1}{2}\|\Phi(X) - y\|^2$
6. K is the number of stars
7. n^2 is the number of pixels we sample (Distributed in a $n \times n$ grid)

Exercise 3

> Obtain an expression for $D\varphi(x)[u]$

If we define $h_i(x) = h(p_i - x)$ for $i \in \{1, \ldots, n^2\}$, then we have

$$\begin{aligned}
D\varphi(x)[u] &= \lim_{t\to 0} \frac{\varphi(x+tu) - \varphi(x)}{t} \\
&= \lim_{t\to 0} \frac{1}{t} \begin{bmatrix} h(p_1 - x - tu) - h(p_1 - x) \\ \vdots \\ h(p_{n^2} - x - tu) - h(p_{n^2} - x) \end{bmatrix} = \\
&= \begin{bmatrix} \lim_{t\to 0} \frac{h(p_1 - x - tu) - h(p_1 - x)}{t} \\ \vdots \\ \lim_{t\to 0} \frac{h(p_{n^2} - x - tu) - h(p_{n^2} - x)}{t} \end{bmatrix} = \\
&= \begin{bmatrix} Dh(p_1 - x)[-u] \\ \vdots \\ Dh(p_{n^2} - x)[-u] \end{bmatrix} = \\
&= \begin{bmatrix} -Dh(p_1 - x)[u] \\ \vdots \\ -Dh(p_{n^2} - x)[u] \end{bmatrix} = \begin{bmatrix} Dh_1(x)[u] \\ \vdots \\ Dh_{n^2}(x)[u] \end{bmatrix}
\end{aligned}$$

It's easy to see that $\nabla h(x) = \frac{-2x}{\sigma^2} \exp\left(-\frac{\|x\|^2}{\sigma^2}\right)$, and substituting $D_{h_i}(x)[u]$ for $\langle \nabla h_i(x), u \rangle$, in our previous expression for $D\varphi(x)[u]$, we observe that thanks to h_i having a gradient, we can express it as

$$D\varphi(x)[u] = \begin{bmatrix} Dh_1(x)[u] \\ \vdots \\ Dh_{n^2}(x)[u] \end{bmatrix} = \begin{bmatrix} \langle \nabla h_1(x), u \rangle \\ \vdots \\ \langle \nabla h_{n^2}(x), u \rangle \end{bmatrix} = A(x)u$$

where

$$A : \mathbb{R}^2 \to \mathbb{R}^{n^2 \times 2} : x \mapsto \begin{bmatrix} \cdots \nabla h_1(x) \cdots \\ \vdots \quad \vdots \quad \vdots \\ \cdots \nabla h_{n^2}(x) \cdots \end{bmatrix} = \begin{bmatrix} \cdots - \nabla h(p_1 - x) \cdots \\ \vdots \quad \vdots \quad \vdots \\ \cdots - \nabla h(p_{n^2} - x) \cdots \end{bmatrix}$$

Exercise 4

Since the adjoint of a matrix $A(x)$ is its transpose, we have that

$$A(x)^* := \begin{bmatrix} \vdots & \cdots & \vdots \\ \nabla h_1(x) & \cdots & \nabla h_{n^2}(x) \\ \vdots & \cdots & \vdots \end{bmatrix} = A(x)^T$$

Exercise 5

Recall how we defined the function f as a function of other functions

- $f(X) = \frac{1}{2} \|\Phi(X) - y\|^2$
- $\Phi(X) = \sum_{k=1}^{K} \varphi(x_k)$
- $\varphi(x) = \begin{bmatrix} h(p_1 - x) \\ \vdots \\ h(p_{n^2} - x) \end{bmatrix} = [h(p_i - x)]_{i=1}^{n^2}$

And finally, the definition of the dot product $\langle A, B \rangle$ for any matrices $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{n \times k}$ $(n, m, k > 0)$

$$\langle A, B \rangle = tr(A^T B)$$

First, we recall that $D_\varphi(x)[u] = A(x)\ u$, where $A(x) \in \mathbb{R}^{n^2 \times 2}$.

With this information, we get the following expression for

$$\begin{aligned} D\Phi(X)[U] &= \sum_{k=1}^{K} D\varphi(x_k)[u_k] = \\ &= \sum_{k=1}^{K} A(x_k)u_k \end{aligned}$$

And thanks to this, we can find a final expression for $D_f(X)(U)$

$$\begin{aligned} Df(X)[U] &= \langle \Phi(X) - y, D\Phi(X)[u_k] \rangle = \\ &= \left\langle \Phi(X) - y, \sum_{k=1}^{K} A(x_k)u_k \right\rangle = \\ &= \sum_{k=1}^{K} \langle \Phi(X) - y, A(x_k)u_k \rangle = \\ &= \sum_{k=1}^{K} \langle A(x_k)^*(\Phi(X) - y), u_k \rangle = \\ &= tr\left(\left([A(x_k)^*(\Phi(X) - y)]_{k=1}^{K} \right)^T U \right) = \\ &= \left\langle [A(x_k)^*(\Phi(X) - y)]_{k=1}^{K}, U \right\rangle \end{aligned}$$

And so, thanks to the definition of gradient, we can identify the gradient of f as

$$\nabla f(X) = [A(x_k)^*(\Phi(X) - y)]_{k=1}^{K} = \begin{bmatrix} A(x_1)^*(\Phi(X) - y) \\ \vdots \\ A(x_K)^*(\Phi(X) - y) \end{bmatrix}$$