Q1.1

1. $\frac{\partial W(x;p)}{\partial p^T}$

$$W(\vec{x};\vec{p}) = \vec{x}+\vec{p} = \begin{pmatrix} x+p_1 \\ y+p_2 \end{pmatrix} = \begin{pmatrix} W_x(x,y) \\ W_y(x,y) \end{pmatrix}$$

$$\frac{\partial W(\vec{x};\vec{p})}{\partial \vec{p}^T} = \begin{pmatrix} \frac{\partial W_x(x,y)}{\partial p_1} & \frac{\partial W_x(x,y)}{\partial p_2} \\ \frac{\partial W_y(x,y)}{\partial p_1} & \frac{\partial W_y(x,y)}{\partial p_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

2. $$\left| I_{t+1}(x'+\Delta p) - I_t(x) \right|^2$$

$$= \left| I_{t+1}(x') + \frac{\partial I_{t+1}(x')}{\partial x'^T} \frac{\partial W(x;p)}{\partial p^T}\Delta p - I_t(x) \right|^2$$

$$= \left| \frac{\partial I_{t+1}(x')}{\partial x'^T} \frac{\partial W(x;p)}{\partial p^T}\Delta p - (I_t(x) - I_{t+1}(x')) \right|^2$$

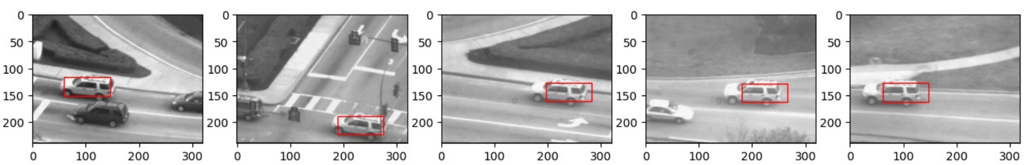
$$A = \begin{bmatrix} \frac{\partial I_{t+1}(x_1)}{\partial x_1^T} \\ \vdots \\ \frac{\partial I_{t+1}(x_D)}{\partial x_D^T} \end{bmatrix}, \quad b = \begin{bmatrix} I_t(x_1) - I_{t+1}(x_1') \\ I_t(x_2) - I_{t+1}(x_2') \\ \vdots \\ I_t(x_D) - I_{t+1}(x_D') \end{bmatrix}$$

as $\frac{\partial W(x;p)}{\partial p^T} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

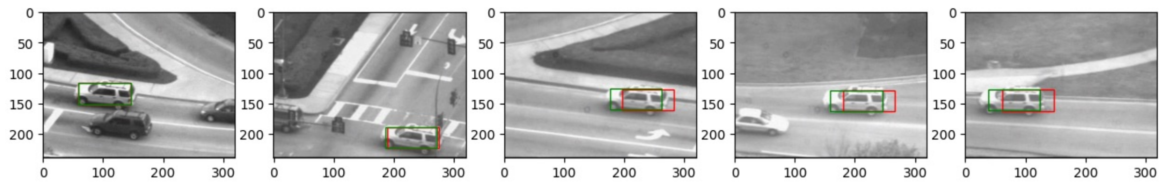
A^TA is invertible.

It's better that both eigenvalues are large and have similar magnitude.

Q1.3



Q1.4



Q2.1

$$I_{t+1}(x) = I_t(x) + w^T \cdot B(x)$$

asume n is the number of pixels in the frame.

I_{t+1} I_t are n×1 vectors and $\{B_k\}_{k=1}^K$ is K×n matrix

let $B = \{B_k\}_{k=1}^K$

$$(I_{t+1} - I_t) = w^T \cdot B$$

$$B_i^T (I_{t+1} - I_t) = B_i^T w^T B = w_i B_i^T \cdot B_i$$

because $B_i^T B_j = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$

$$w_i = B_i^T (I_{t+1} - I_t)$$

$$w = (B_1^T \quad B_2^T \quad \dots \quad B_K^T)(I_{t+1} - I_t)$$

$$\| B^\perp (A\Delta p - b) \|_2^2$$

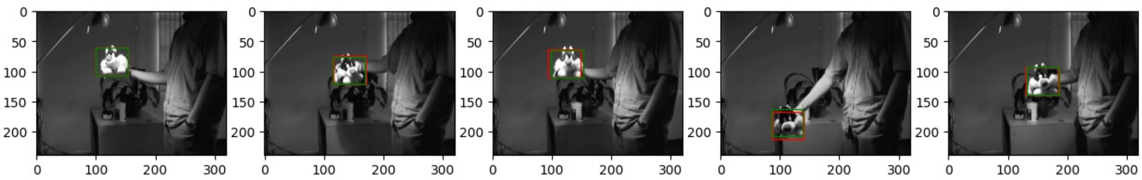
$$= \| A\Delta p - b - BB^T \|$$

$$\| A\Delta p - b \|^2_{span(B^\perp)} = \| A\Delta p - b - BB^T(A\Delta p - b) \|_2^2$$

$$= \| (A - BB^T A)\Delta p - (b - BB^T b) \|_2^2$$

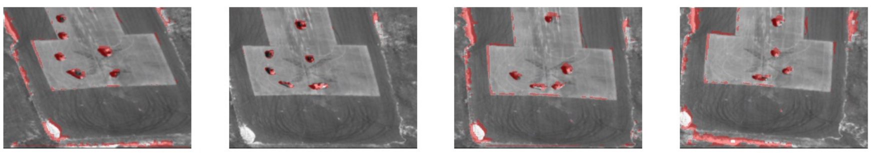
$$= \| (I - BB^T) A\Delta p - (I - BB^T) b \|_2^2$$

Q2.3

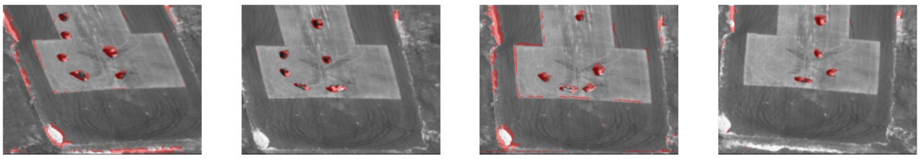


The difference of performance is relatively small.

Q3.3



Q4.1



The gradient of T and Jacobian can be precomputed. So we only need to compute affine of I in each iteration.

$$\text{for } \nabla T(W(x;0)) = \nabla T$$

Q4.2 $\arg\min_{g} \frac{1}{2}\|y - x^T g\|_2^2 + \frac{\lambda}{2}\|g\|_2^2$

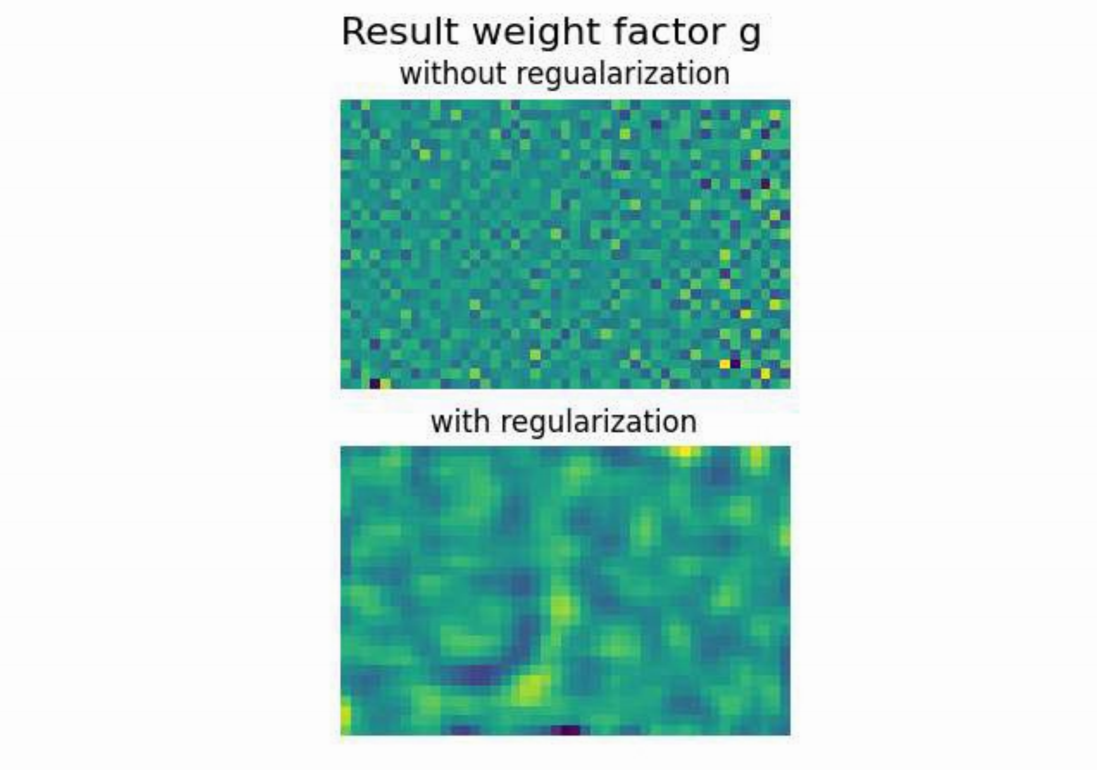
$$
\begin{aligned}
L &= \frac{1}{2}\|y - x^T g\|_2^2 + \frac{\lambda}{2}\|g\|^2 \\
&= \frac{1}{2}(y - x^T g)^T (y - x^T g) + \frac{\lambda}{2} g^T g \\
&= \frac{1}{2}(y^T - g^T x)(y - x^T g) + \frac{\lambda}{2} g^T g \\
&= \frac{1}{2} y^T y - y^T x^T g + \frac{1}{2} g^T x x^T g + \frac{\lambda}{2} g^T g
\end{aligned}
$$

$$\frac{\partial L}{\partial g} = -y^T x^T + g^T(xx^T + \lambda I) = 0$$

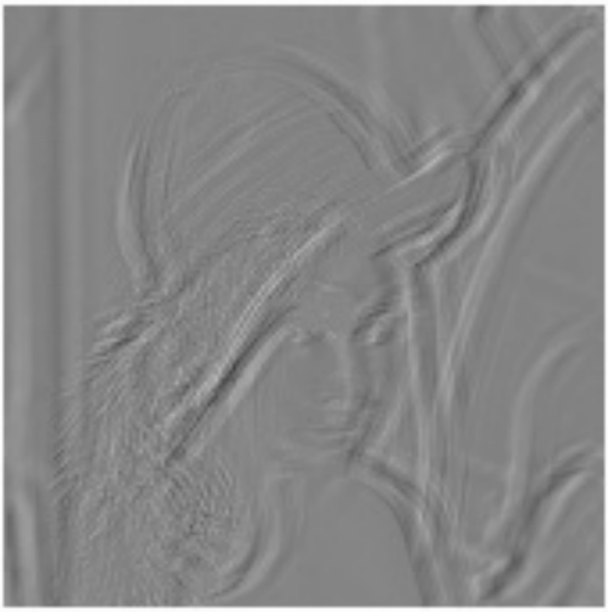
$$-xy + (xx^T + \lambda I)g = 0$$

$$
\begin{aligned}
g &= (xx^T + \lambda I)^{-1} x y \\
&= (S + \lambda I)^{-1} X y
\end{aligned}
$$

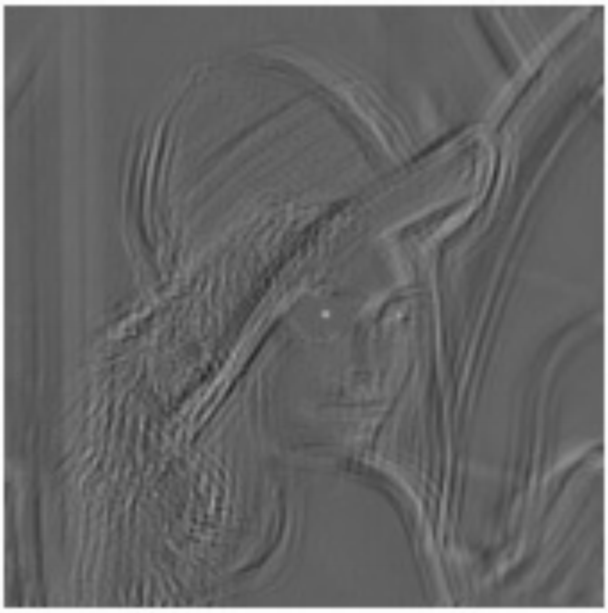
Q4.3



without regualarization

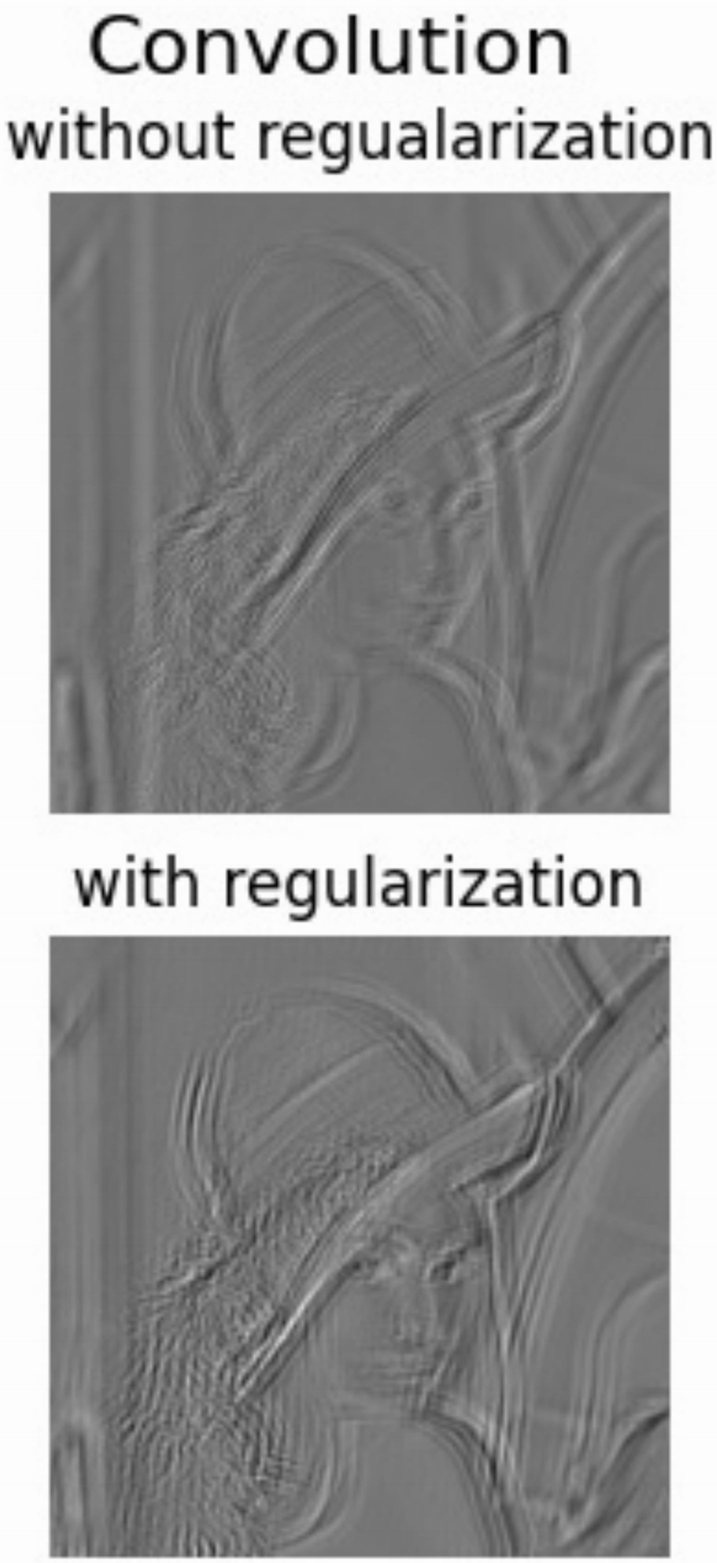


with regularization



$\lambda=1$ work the bese. Because $\|g\|_2^2$ term prevent the model from being too complex. And then no feautures will dominate the predicting result.

Q4.4



Because convolution is different from convolution When g[::-1, ::-1] is performed, they should be equivalent