# ADMM algorithm

LongZhang *

July 8, 2025

# 1 Imaging

We know that imaging (F) by an optical imaging system equals the ideal image (X) convolved with the point spread function (K):

$$F = K * X + n \tag{1}$$

where $n$ denotes random noise.

## 1.1 Why ADMM?

Why Direct Inverse Filtering Fails in Practical Non-Blind Image Restoration?

Theoretical Perfect Restoration: In theory, for a Linear Space-Invariant (LSI) system without noise, the perfect restoration formula is:

Fourier Transform → Divide by OTF → Inverse Fourier Transform

However, inverse filtering nearly always fails in practical non-blind image restoration due to:

## 1.2 OTF Zero Points:

1. The Optical Transfer Function (OTF, Fourier transform of PSF) contains frequency components where $OTF(\omega) = 0$.
2. Division by OTF at these points becomes infinite (1/0).
3. Even near-zero OTF values cause extreme amplification of corresponding frequency components.

## 1.3 Noise Amplification:

1. Real images always contain noise.
2. Noise predominantly resides in high-frequency regions.
3. OTF values are typically very small at high frequencies.
4. Division by small OTF values amplifies noise catastrophically.

## 1.4 PSF Truncation and Modeling Errors:

1. Measured PSFs are finite-sized, truncated approximations.
2. Truncation introduces artifacts in the OTF representation.
3. Accurate PSF modeling is challenging and error-prone.

## 1.5 Ill-Posed Nature:

1. Due to OTF zeros or near-zeros, restoration becomes ill-posed.
2. Small perturbations lead to massive variations in output.

# 2 Theory

## 2.1 Problem Formulation

**Objective function:**

$$\begin{aligned}\min_X \sum_i \|D_i X\|_2 \\ + \frac{\mu}{2}\|K * X - F\|_F^2\end{aligned} \tag{2}$$

where $D_i X$ denotes the gradient of image $X$ at pixel $i$, $\|\cdot\|_2$ is the vector 2-norm, and $\|\cdot\|_F$ is the Frobenius norm.

*Computational optical imaging; Picture processing

## 2.2 ADMM Formulation

**Introduce auxiliary variables:** Let $Y = (Y_1, Y_2)$ replace $DX = (D_x X, D_y X)$:

$$\min_{X,Y} \quad \|Y\|_{2,1} + \frac{\mu}{2}\|K * X - F\|_F^2 \tag{3}$$

$$\text{s.t.} \quad Y = DX \tag{4}$$

where $\|Y\|_{2,1} = \sum_{i,j} \sqrt{(Y_1)_{i,j}^2 + (Y_2)_{i,j}^2}$ is the isotropic TV norm.

**Augmented Lagrangian:** Introduce Lagrange multiplier $\Lambda = (\Lambda_1, \Lambda_2)$:

$$\begin{aligned}\mathcal{L}(X, Y, \Lambda) = \|Y\|_{2,1} + \frac{\mu}{2}\|K * X - F\|_F^2 \\ + \langle \Lambda, DX - Y\rangle \\ + \frac{\beta}{2}\|DX - Y\|_F^2\end{aligned} \tag{5}$$

where $\beta > 0$ is the penalty parameter.

## 2.3 ADMM Algorithm

**(a) Update Y (Shrinkage step):** Fix $X$ and $\Lambda$, solve:

$$\min_Y \|Y\|_{2,1} + \frac{\beta}{2}\|DX - Y + \beta^{-1}\Lambda\|_F^2 \tag{6}$$

Let $Z = DX + \beta^{-1}\Lambda$. The problem becomes:

$$\min_Y \|Y\|_{2,1} + \frac{\beta}{2}\|Y - Z\|_F^2 \tag{7}$$

**Derivation of the Shrinkage Solution:** The optimization problem can be decomposed per pixel since both terms are separable. For each pixel $(i, j)$, we solve:

$$\min_{v\in\mathbb{R}^2} \|v\|_2 + \frac{\beta}{2}\|v - z\|_2^2 \tag{8}$$

where $v = Y_{i,j}$, $z = Z_{i,j}$.

This is a proximal operator for the $\ell_2$-norm. To solve it, we consider the subdifferential of the objective function $f(v) = \|v\|_2 + \frac{\beta}{2}\|v - z\|_2^2$:

- **Case 1:** $\|z\|_2 > \frac{1}{\beta}$

  The solution occurs when the gradient is zero:

$$\frac{v}{\|v\|_2} + \beta(v - z) = 0$$

$$v = \left(1 - \frac{1}{\beta\|z\|_2}\right) z$$

- **Case 2:** $\|z\|_2 \le \frac{1}{\beta}$

  The optimal solution is $v = 0$, as verified by the subdifferential condition:

$$0 \in \partial f(0) = \{g - \beta z \mid \|g\|_2 \le 1\}$$

$$\|\beta z\|_2 \le 1$$

Combining both cases, we obtain the vector shrinkage operator:

$$v = \frac{z}{\|z\|_2} \max\left(\|z\|_2 - \frac{1}{\beta}, 0\right) \tag{9}$$

Therefore, the solution for the entire image is:

$$Y_{i,j} = \frac{Z_{i,j}}{\|Z_{i,j}\|_2} \max\left(\|Z_{i,j}\|_2 - \frac{1}{\beta}, 0\right) \tag{10}$$

where $Z_{i,j} = ((Z_1)_{i,j}, (Z_2)_{i,j}) \in \mathbb{R}^2$.

**(b) Update X (Linear subproblem):** Fix $Y$ and $\Lambda$, solve:

$$\min_X \frac{\mu}{2}\|K * X - F\|_F^2 + \frac{\beta}{2}\|DX - Y + \beta^{-1}\Lambda\|_F^2 \tag{11}$$

Optimality condition:

$$\mu K^*(K * X - F) + \beta D^*(DX - Y + \beta^{-1}\Lambda) = 0 \tag{12}$$

Rearranged form:

$$(\mu K^* K + \beta D^* D)X = \mu K^* F + \beta D^*(Y - \beta^{-1}\Lambda) \tag{13}$$

Frequency domain solution:

$$X = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mu K^* F + \beta D^*(Y - \beta^{-1}\Lambda))}{\mu|\mathcal{F}(K)|^2 + \beta(|\mathcal{F}(D_x)|^2 + |\mathcal{F}(D_y)|^2)}\right) \tag{14}$$

Numerator implementation:

$$\text{Numerator} = \mu K^\top F - D^*(\Lambda) + \beta D^*(Y) \tag{15}$$

**(c) Update multiplier:**

$$\Lambda \leftarrow \Lambda - \gamma\beta(DX - Y) \tag{16}$$

where $\gamma \in (0, 1.618]$ is a relaxation parameter.

## 2.4 Precomputed Frequency Domain Constants

$\texttt{H} = \mathcal{F}(H, \text{size}(F))$

$\texttt{K*F} = \mathcal{F}^{-1}\left(\overline{\texttt{H}} \cdot \mathcal{F}(F)\right)$

$|\mathcal{F}(D)|^2 = |\mathcal{F}([1, -1], \text{sizeF})|^2 + |\mathcal{F}([1; -1], \text{sizeF})|^2$

### 2.5 Finite Difference Operators

**Forward difference:**

$$(D_x U)_{i,j} = \begin{cases} U_{i,j+1} - U_{i,j} & j < n \\ U_{i,1} - U_{i,n} & j = n \end{cases}$$

$$(D_y U)_{i,j} = \begin{cases} U_{i+1,j} - U_{i,j} & i < m \\ U_{1,j} - U_{m,j} & i = m \end{cases}$$

**Negative divergence ($D^* = -$div):**

$$(D^*(V))_{i,j} = (V^x_{i,j} - V^x_{i,j-1}) + (V^y_{i,j} - V^y_{i-1,j}) \tag{17}$$

with circular boundary conditions:

$$V^x_{i,0} = V^x_{i,n}, \quad V^x_{i,n+1} = V^x_{i,1}$$
$$V^y_{0,j} = V^y_{m,j}, \quad V^y_{m+1,j} = V^y_{1,j}$$

### 2.6 Algorithm

**Algorithm 1** ADMM for Image Restoration

**Input:** Blurred image $F$, kernel $K$, parameters $\mu, \beta, \gamma$

**Initialize:** Restored image $X$

Initialize $X^0 = F$, $\Lambda^0 = 0$

**for** $k = 0, 1, 2, \dots$ **do**

  Update $Y^{k+1}$ via shrinkage: Eq.(12)

  Update $X^{k+1}$ via frequency domain: Eq.(16)

  **if** $\|X^{k+1} - X^k\| / \|X^k\| < \epsilon$ **then**

    **break**

  **end if**

  Update multiplier: $\Lambda^{k+1} = \Lambda^k - \gamma\beta(DX^{k+1} - Y^{k+1})$

**end for**

**Output:** Restored image $X$

### 2.7 TV/L1 Model: Robust Restoration with Impulse Noise

**Problem Formulation:** The TV/L1 model addresses impulse noise (salt-and-pepper, random-valued) by replacing the L2 fidelity term with an L1 norm:

$$\min_X \sum_i \|D_i X\|_2 + \mu \|K * X - F\|_1 \tag{18}$$

where $\|\cdot\|_1$ denotes the element-wise L1 norm. This formulation is robust to outliers in the observed image $F$.

**Key Differences from TV/L2:**

1. **Fidelity Term:** L1 norm ($\|\cdot\|_1$) replaces L2 norm ($\|\cdot\|_F^2$)
2. **Regularization Parameter:** $\mu$ scales the fidelity term directly (no 1/2 factor)
3. **Noise Assumption:** Designed for impulse noise instead of Gaussian noise

### 2.8 ADMM Formulation with Three Splits

To handle the non-differentiable L1 term, we introduce **three auxiliary variables**:

$$Y_1 = D_x X \quad \text{(horizontal gradients)} \tag{19}$$

$$Y_2 = D_y X \quad \text{(vertical gradients)} \tag{20}$$

$$Z = K * X - F \quad \text{(residuals)} \tag{21}$$

The constrained problem becomes:

$$\min_{X, Y_1, Y_2, Z} \quad \|Y_1\|_2 + \|Y_2\|_2 + \mu \|Z\|_1 \tag{22}$$

$$\text{s.t.} \quad Y_1 = D_x X, \quad Y_2 = D_y X, \quad Z = K * X - F \tag{23}$$

where $\|Y_1\|_2 = \sum_{i,j} \sqrt{(Y_1)^2_{i,j}}$, similarly for $Y_2$.

### 2.9 Augmented Lagrangian

We introduce Lagrange multipliers $\Lambda_1, \Lambda_2, \Lambda_3$ and penalty parameters $\beta_1, \beta_2$:

$$\begin{aligned} \mathcal{L}(X, Y_1, Y_2, Z, \Lambda_1, \Lambda_2, \Lambda_3) = \\ \|Y_1\|_2 + \|Y_2\|_2 + \mu \|Z\|_1 \\ + \langle \Lambda_1, D_x X - Y_1 \rangle + \frac{\beta_1}{2} \|D_x X - Y_1\|_F^2 \\ + \langle \Lambda_2, D_y X - Y_2 \rangle + \frac{\beta_1}{2} \|D_y X - Y_2\|_F^2 \\ + \langle \Lambda_3, (K * X - F) - Z \rangle + \frac{\beta_2}{2} \|K * X - F - Z\|_F^2 \end{aligned} \tag{24}$$

### 2.10 ADMM Updates

The algorithm alternates between updating variables:

**(a) Update $Y_1$ and $Y_2$ (Group Shrinkage):** For each gradient direction $k \in \{1, 2\}$:

$$Y_k^{t+1} = \arg\min_{Y_k} \|Y_k\|_2 + \frac{\beta_1}{2}\|D_k X^t - Y_k + \beta_1^{-1}\Lambda_k^t\|_F^2 \tag{25}$$

Let $V_k = D_k X^t + \beta_1^{-1}\Lambda_k^t$. The closed-form solution is:

$$Y_k^{t+1} = \max\left(\|V_k\|_2 - \frac{1}{\beta_1}, 0\right)\frac{V_k}{\|V_k\|_2} \tag{26}$$

with $Y_k^{t+1} = 0$ when $\|V_k\|_2 = 0$.

**(b) Update $Z$ (Element-wise Soft Thresholding):**

$$Z^{t+1} = \arg\min_Z \mu\|Z\|_1 + \frac{\beta_2}{2}\|K*X^t - F - Z + \beta_2^{-1}\Lambda_3^t\|_F^2 \tag{27}$$

Let $R = K * X^t - F + \beta_2^{-1}\Lambda_3^t$. The solution is:

$$Z^{t+1} = \text{sign}(R) \odot \max\left(|R| - \frac{\mu}{\beta_2}, 0\right) \tag{28}$$

where $\odot$ denotes element-wise multiplication.

**(c) Update $X$ (Quadratic Subproblem):**

$$\begin{aligned} X^{t+1} = \arg\min_X & \frac{\beta_1}{2}\|D_x X - Y_1^{t+1} + \beta_1^{-1}\Lambda_1^t\|_F^2 \\ &+ \frac{\beta_1}{2}\|D_y X - Y_2^{t+1} + \beta_1^{-1}\Lambda_2^t\|_F^2 \\ &+ \frac{\beta_2}{2}\|K * X - F - Z^{t+1} + \beta_2^{-1}\Lambda_3^t\|_F^2 \end{aligned} \tag{29}$$

Define:

$$W_1 = Y_1^{t+1} - \beta_1^{-1}\Lambda_1^t \tag{30}$$

$$W_2 = Y_2^{t+1} - \beta_1^{-1}\Lambda_2^t \tag{31}$$

$$W_3 = F + Z^{t+1} - \beta_2^{-1}\Lambda_3^t \tag{32}$$

The optimality condition is:

$$\beta_1(D_x^T D_x + D_y^T D_y)X + \beta_2 K^T K * X = \beta_1(D_x^T W_1 + D_y^T W_2) + \beta_2 K^T * W_3 \tag{33}$$

In the frequency domain (assuming periodic boundaries):

$$X^{t+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}\left(\beta_1(D_x^T W_1 + D_y^T W_2) + \beta_2 K^T * W_3\right)}{\beta_1\mathcal{F}(L) + \beta_2|\mathcal{F}(K)|^2}\right) \tag{34}$$

where $L = D_x^T D_x + D_y^T D_y$ is the discrete Laplacian operator.

**(d) Update Multipliers:**

$$\Lambda_1^{t+1} = \Lambda_1^t - \gamma\beta_1(D_x X^{t+1} - Y_1^{t+1}) \tag{35}$$

$$\Lambda_2^{t+1} = \Lambda_2^t - \gamma\beta_1(D_y X^{t+1} - Y_2^{t+1}) \tag{36}$$

$$\Lambda_3^{t+1} = \Lambda_3^t - \gamma\beta_2((KX^{t+1} - F) - Z^{t+1}) \tag{37}$$

with $\gamma \in (0, (1+\sqrt{5})/2]$ typically set to 1.618 for acceleration.

## 2.11 Implementation Notes

- **Variable Initialization:** $X^0 = F$, $Y_1^0 = D_x F$, $Y_2^0 = D_y F$, $Z^0 = K * F - F$, $\Lambda_i^0 = 0$
- **Parameter Selection:** $\beta_1 \approx 0.1\mu$, $\beta_2 \approx \mu$, $\mu$ depends on noise level
- **Convergence:** Stop when $\|X^{t+1} - X^t\|_F / \|X^{t+1}\|_F < \epsilon$ ($\epsilon \sim 10^{-4}$)

## 2.12 Comparison with TV/L2 ADMM

Table 1: TV/L1 vs TV/L2 ADMM Comparison

| **Characteristic** | **TV/L1** | **TV/L2** |
|---|---|---|
| Fidelity Term | $\mu\|KX - F\|_1$ | $\frac{\mu}{2}\|KX - F\|_F^2$ |
| Auxiliary Variables | 3 ($Y_1, Y_2, Z$) | 1 ($Y$) |
| Z-update | Soft thresholding | Not needed |
| Gradient Update | Group shrinkage | Group shrinkage |
| X-update Complexity | Higher (3 terms) | Lower (2 terms) |
| Noise Robustness | Impulse noise | Gaussian noise |