Plug and Play ADMM for Hyperspectral Image

Zeqiang Lai

June 1, 2021

Contents

1 Introduction

Image restoration problem can be formulated as an optimization problem:

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}) + \lambda g(\boldsymbol{x}) \tag{1}$$

where $f(x)$ is a data fidelity term and $g(x)$ is a regularization term.

There are various algorithms could be used to solve this optimization problem, among which the Alternating Direction Method of Multipliers(ADMM) is most popular.

1.1 ADMM

ADMM algorithm introduce a new variable v to decouple the data fidelity term and regularization term:

$$(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{v}}) = \underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}) + \lambda g(\boldsymbol{v}) \quad \text{subject to } \boldsymbol{x} = \boldsymbol{v} \tag{2}$$

After that, the augmented Lagrangian method is used to convert it into an unconstrained problem, whose objective can be describe as:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{u}) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{v}) + \boldsymbol{u}^T(\boldsymbol{x} - \boldsymbol{v}) + \frac{\rho}{2}\|\boldsymbol{x} - \boldsymbol{v}\|^2 \tag{3}$$

ADMM optimizes x, v and u alternately, and it is shown that processing in this way converges to the solution of (2) under some assumption.

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmin}} \quad f(\boldsymbol{x}) + \tfrac{\rho}{2}\left\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(k)}\right\|^2, \quad \tilde{\boldsymbol{x}}^{(k)} \equiv \boldsymbol{v}^{(k)} - \tilde{\boldsymbol{u}}^{(k)} \\
\boldsymbol{v}^{(k+1)} &= \underset{\boldsymbol{v} \in \mathbb{R}^n}{\operatorname{argmin}} \quad \lambda g(\boldsymbol{v}) + \tfrac{\rho}{2}\left\|\boldsymbol{v} - \tilde{\boldsymbol{v}}^{(k)}\right\|^2, \quad \tilde{\boldsymbol{v}}^{(k)} \equiv \boldsymbol{x}^{(k+1)} + \bar{\boldsymbol{u}}^{(k)} \\
\bar{\boldsymbol{u}}^{(k+1)} &= \bar{\boldsymbol{u}}^{(k)} + \left(\boldsymbol{x}^{(k+1)} - \boldsymbol{v}^{(k+1)}\right), \quad \bar{\boldsymbol{u}}^{(k)} \equiv (1/\rho)\boldsymbol{u}^{(k)}
\end{aligned} \tag{4}$$

1.2 Plug and Play ADMM

If we define $\sigma \equiv \sqrt{\lambda\rho}$, we could rewrite the second equation in (4) as:

$$\boldsymbol{v}^{(k+1)} = \underset{\boldsymbol{v} \in \mathbb{R}^n}{\operatorname{argmin}} \quad g(\boldsymbol{v}) + \frac{1}{2\sigma^2}\left\|\boldsymbol{v} - \tilde{\boldsymbol{v}}^{(k)}\right\|^2 \tag{5}$$

Intuitively, (5) can be considered as a denosing problem, where σ is the noisy level, $\boldsymbol{v}$ is the "clean" image, and $\tilde{\boldsymbol{v}}^k$ is the corresponding "noisy" one.

Further, (5) can be solved with an off-the-shelf image denoising algorithm:

$$\boldsymbol{v}^{(k+1)} = \mathcal{D}_\sigma\left(\tilde{\boldsymbol{v}}^{(k)}\right) \tag{6}$$

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmin}} \quad f(\boldsymbol{x}) + \tfrac{\rho}{2}\left\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(k)}\right\|^2, \quad \tilde{\boldsymbol{x}}^{(k)} \equiv \boldsymbol{v}^{(k)} - \tilde{\boldsymbol{u}}^{(k)} \\
\boldsymbol{v}^{(k+1)} &= \mathcal{D}_\sigma\left(\tilde{\boldsymbol{v}}^{(k)}\right), \quad \tilde{\boldsymbol{v}}^{(k)} \equiv \boldsymbol{x}^{(k+1)} + \bar{\boldsymbol{u}}^{(k)} \\
\bar{\boldsymbol{u}}^{(k+1)} &= \bar{\boldsymbol{u}}^{(k)} + \left(\boldsymbol{x}^{(k+1)} - \boldsymbol{v}^{(k+1)}\right), \quad \bar{\boldsymbol{u}}^{(k)} \equiv (1/\rho)\boldsymbol{u}^{(k)}
\end{aligned} \tag{7}$$

2 Application

As it is indicated by Equation 7, the only difference of different application of Plug and Play ADMM lies in $f(x)$.

In the image restoration problem, $f(x)$ has the following general form:

$$f(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||^2 \tag{8}$$

where $\boldsymbol{A}$ is a transformation matrix.

For now, we have to optimize the follow objective:

$$\boldsymbol{x}^{(k+1)} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}} \quad \frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||^2 + \frac{\rho}{2}\left\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(k)}\right\|^2 \tag{9}$$

With $\tilde{\boldsymbol{x}}^{(k)} \equiv \boldsymbol{v}^{(k)} - \tilde{\boldsymbol{u}}^{(k)}$ fixed, this is equivalent to the least-squares problem

$$\min_x \left\| \begin{bmatrix} A \\ \sqrt{\rho} I \end{bmatrix} x - \begin{bmatrix} y \\ \sqrt{\rho}\tilde{\boldsymbol{x}}^{(k)} \end{bmatrix} \right\|^2 \tag{10}$$

This problem has a closed-form solution:

$$\begin{aligned} x^{(k+1)} &= \left(A^T A + \rho I\right)^{-1} \left[A^T, \sqrt{\rho} I\right] \begin{bmatrix} y \\ \sqrt{\rho}\tilde{x}^{(k)} \end{bmatrix} \\ &= \left(A^T A + \rho I\right)^{-1} \left(A^T y + \rho \tilde{x}^{(k)}\right) \end{aligned} \tag{11}$$

2.1 Deblur

In Debluring task:

$$f(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}||^2 \tag{12}$$

where H is a circulant matrix denoting the blur.

$$x^{(k+1)} = \left(H^T H + \rho I\right)^{-1} \left(H^T y + \rho \tilde{x}^{(k)}\right) \tag{13}$$

2.2 Single Image Super Resolution

In Super resolution task:

$$f(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{S}\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}||^2 \tag{14}$$

where H is a circulant matrix denoting the blur. S is a binary matrix denoting the K-fold downsampling,

$$\boldsymbol{x} = \rho^{-1}\boldsymbol{b} - \rho^{-1}\boldsymbol{G}^T \left(\mathcal{F}^{-1} \left\{ \frac{\mathcal{F}(\boldsymbol{G}\boldsymbol{b})}{\left|\mathcal{F}\left(\widetilde{h}_0\right)\right|^2 + \rho} \right\} \right) \tag{15}$$

where $\boldsymbol{G} = \boldsymbol{S}\boldsymbol{H}, \boldsymbol{b} = \boldsymbol{G}^T \boldsymbol{y} + \rho\widetilde{\boldsymbol{x}}$, $\tilde{h}_0$ is the 0th polyphase component of the filter HH^T.

2.3 Multi Image Super Resolution

$$\begin{aligned}\mathcal{L}(\boldsymbol{x},\boldsymbol{v},\boldsymbol{u}) =& \frac{1}{2}||\boldsymbol{T}\boldsymbol{x}-\boldsymbol{z}||^2 \\ &+\frac{1}{2}||\boldsymbol{S}\boldsymbol{H}\boldsymbol{w}-\boldsymbol{y}||^2+\boldsymbol{m}^T(\boldsymbol{x}-\boldsymbol{w})+\frac{\rho}{2}\|\boldsymbol{x}-\boldsymbol{w}\|^2 \\ &+\lambda g(\boldsymbol{v})+\boldsymbol{u}^T(\boldsymbol{x}-\boldsymbol{v})+\frac{\rho}{2}\|\boldsymbol{x}-\boldsymbol{v}\|^2\end{aligned} \tag{16}$$

$$\begin{aligned}
&\boldsymbol{x}^{(k+1)}=\underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}} \quad \tfrac{1}{2}||\boldsymbol{T}\boldsymbol{x}-\boldsymbol{z}||^2+\tfrac{\rho}{2}\left\|\boldsymbol{x}-\tilde{\boldsymbol{x}}^{(k)}\right\|^2+\tfrac{\rho}{2}\left\|\boldsymbol{x}-\tilde{\boldsymbol{x_2}}^{(k)}\right\|^2, \quad \tilde{\boldsymbol{x}}^{(k)}\equiv\boldsymbol{v}^{(k)}-\tilde{\boldsymbol{u}}^{(k)} \\
&\boldsymbol{w}^{(k+1)}=\underset{\boldsymbol{v}\in\mathbb{R}^n}{\operatorname{argmin}} \quad \tfrac{1}{2}||\boldsymbol{S}\boldsymbol{H}\boldsymbol{w}-\boldsymbol{y}||^2+\tfrac{\rho}{2}\left\|\boldsymbol{w}-\widetilde{\boldsymbol{w}}^{(k)}\right\|^2, \quad \tilde{\boldsymbol{v}}^{(k)}\equiv\boldsymbol{x}^{(k+1)}+\bar{\boldsymbol{m}}^{(k)} \\
&\boldsymbol{v}^{(k+1)}=\underset{\boldsymbol{v}\in\mathbb{R}^n}{\operatorname{argmin}} \quad \lambda g(\boldsymbol{v})+\tfrac{\rho}{2}\left\|\boldsymbol{v}-\widetilde{\boldsymbol{v}}^{(k)}\right\|^2, \quad \tilde{\boldsymbol{v}}^{(k)}\equiv\boldsymbol{x}^{(k+1)}+\bar{\boldsymbol{u}}^{(k)} \\
&\bar{\boldsymbol{u}}^{(k+1)}=\bar{\boldsymbol{u}}^{(k)}+\left(\boldsymbol{x}^{(k+1)}-\boldsymbol{v}^{(k+1)}\right), \quad \bar{\boldsymbol{u}}^{(k)}\equiv(1/\rho)\boldsymbol{u}^{(k)} \\
&\bar{\boldsymbol{m}}^{(k+1)}=\bar{\boldsymbol{m}}^{(k)}+\left(\boldsymbol{x}^{(k+1)}-\boldsymbol{w}^{(k+1)}\right), \quad \bar{\boldsymbol{u}}^{(k)}\equiv(1/\rho)\boldsymbol{u}^{(k)}
\end{aligned} \tag{17}$$

2.4 Inpainting

In Inpainting task:

$$f(\boldsymbol{x})=\frac{1}{2}||\boldsymbol{S}\boldsymbol{x}-\boldsymbol{y}||^2 \tag{18}$$

where S is a diagonal masking matrix.

$$x^{(k+1)}=\left(S^TS+\rho I\right)^{-1}\left(S^Ty+\rho\tilde{x}^{(k)}\right) \tag{19}$$

S^TS is also diagonal, so $(S^TS+\rho I)$ is diagonal, and matrix inversion $(S^TS+\rho I)^{-1}$ can be implemented as element-wise division.

S is diagonal, so $S^T=S$. $S^Ty=Sy$ is the masking process and can be implemented as element-wise multiplication.

2.5 Compress Sensing

In Compress Sensing task:

$$f(\boldsymbol{x})=\frac{1}{2}||\boldsymbol{\Phi}\boldsymbol{x}-\boldsymbol{y}||^2 \tag{20}$$

where $\boldsymbol{\Phi}\in R^{n\times nB}$ is the sensing matrix, $x\in R^{nB}$ is the origin signal, $y\in R^n$ is the compressed signal.

$$x^{(k+1)}=\left(\Phi^T\Phi+\rho I\right)^{-1}\left(\Phi^Ty+\rho\tilde{x}^{(k)}\right) \tag{21}$$

$\Phi^T\Phi+\rho I$ is of size $nB\times nB$, computing its inverse directly is unacceptable.

$$\left(\boldsymbol{\Phi}^\top\boldsymbol{\Phi}+\rho\boldsymbol{I}\right)^{-1}=\rho^{-1}\boldsymbol{I}-\rho^{-1}\boldsymbol{\Phi}^\top\left(\boldsymbol{I}+\boldsymbol{\Phi}\rho^{-1}\boldsymbol{\Phi}^\top\right)^{-1}\boldsymbol{\Phi}\rho^{-1} \tag{22}$$

Plug 20 into 19, we have

$$\begin{aligned}\boldsymbol{x}^{(k+1)}=&\frac{\left[\boldsymbol{\Phi}^\top\boldsymbol{y}+\rho\tilde{\boldsymbol{x}}^{(k)}\right]}{\rho}-\frac{\boldsymbol{\Phi}^\top\left(\boldsymbol{I}+\boldsymbol{\Phi}\rho^{-1}\boldsymbol{\Phi}^\top\right)^{-1}\boldsymbol{\Phi}\boldsymbol{\Phi}^\top\boldsymbol{y}}{\rho^2} \\ &-\frac{\boldsymbol{\Phi}^\top\left(\boldsymbol{I}+\boldsymbol{\Phi}\rho^{-1}\boldsymbol{\Phi}^\top\right)^{-1}\boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)}}{\rho}\end{aligned} \tag{23}$$

$\Phi\Phi^T$ is a diagonal matrix. Let

$$\Phi\Phi^\top \overset{\text{def}}{=} \operatorname{diag}\{\psi_1, \ldots, \psi_n\} \tag{24}$$

we have:

$$\left(I + \Phi\rho^{-1}\boldsymbol{\Phi}^\top\right)^{-1} = \operatorname{diag}\left\{\frac{\rho}{\rho+\psi_1}, \ldots, \frac{\rho}{\rho+\psi_n}\right\} \tag{25}$$

$$\left(I + \boldsymbol{\Phi}\rho^{-1}\boldsymbol{\Phi}^\top\right)^{-1} \boldsymbol{\Phi}\Phi^\top = \operatorname{diag}\left\{\frac{\rho\psi_1}{\rho+\psi_1}, \ldots, \frac{\rho\psi_n}{\rho+\psi_n}\right\} \tag{26}$$

and

$$\begin{aligned}
\theta =& \frac{1}{\rho}\boldsymbol{\Phi}^\top \boldsymbol{y} + \tilde{\boldsymbol{x}}^{(k)} - \frac{1}{\rho}\boldsymbol{\Phi}^T \operatorname{diag}\left\{\frac{\psi_1}{\rho+\psi_1}, \ldots, \frac{\psi_n}{\rho+\psi_n}\right\} \boldsymbol{y} - \frac{1}{\rho}\boldsymbol{\Phi}^T \operatorname{diag}\left\{\frac{\rho}{\rho+\psi_1}, \ldots, \frac{\rho}{\rho+\psi_n}\right\} \boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)} \\
=& \tilde{\boldsymbol{x}}^{(k)} + \frac{1}{\rho}\boldsymbol{\Phi}^\top \boldsymbol{y} - \frac{1}{\rho}\boldsymbol{\Phi}^\top \left[\frac{y_1\psi_1 + \rho[\boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)}]_1}{\rho+\psi_1}, \ldots, \frac{y_n\psi_n + \rho[\boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)}]_n}{\rho+\psi_n}\right]^\top \\
=& \tilde{\boldsymbol{x}}^{(k)} + \frac{1}{\rho}\boldsymbol{\Phi}^\top \left[\frac{y_1(\rho+\psi_1) - y_1\psi_1 - \rho[\boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)}]_1}{\rho+\psi_1}, \ldots, \frac{y_n(\rho+\psi_n) - y_n\psi_n - \rho[\boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)}]_n}{\rho+\psi_n}\right]^\top \\
=& \boldsymbol{x}^{(k)} + \boldsymbol{\Phi}^\top \left[\frac{y_1 - [\boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)}]_1}{\rho+\psi_1}, \ldots, \frac{y_n - [\boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)}]_n}{\rho+\psi_n}\right]^\top \\
=& \boldsymbol{x}^{(k)} + \boldsymbol{\Phi}^\top \left[(y - \boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{(k)}) \operatorname{diag}\{\rho+\psi_1, \ldots, \rho+\psi_n\}^{-1}\right]
\end{aligned} \tag{27}$$

3 TV Regularized Deep Image Prior

If we add an additional Total Variation Prior, we get the following objective:

$$\text{minimize} \quad \frac{1}{2}\|Ax-y\|_2^2 + \phi g(x) + \lambda\|D_r x\|_1 + \lambda\|D_c x\|_1 \tag{28}$$

Separate different prior by introducing three new variables:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax-y\|_2^2 + \lambda\|z_r\|_1 + \lambda\|z_c\|_1 + \phi g(v) \\ \text{subject to} & D_r x - z_r = 0 \\ \text{subject to} & D_c x - z_c = 0 \\ \text{subject to} & x - v = 0 \end{array} \tag{29}$$

Use augmented lagrangian to eliminate constraints:

$$\begin{aligned} L_\rho(x, z_r, \nu_r, z_c, \nu_c, v) = \frac{1}{2}\|Ax-y\|_2^2 + \lambda\|z_r\|_1 + \nu_r^T(D_r x - z_r) + \frac{\rho}{2}\|D_r x - z_r\|_2^2 \\ + \lambda\|z_c\|_1 + \nu_c^T(D_c x - z_c) + \frac{\rho}{2}\|D_c x - z_c\|_2^2 \\ + \phi g(v) + u^T(x - v) + \frac{\Phi}{2}\|x - v\|_2^2 \end{aligned} \tag{30}$$

Let $\mu_r = \nu_r/\rho$, $\mu_c = \nu_c/\rho$, $\mu = u/\Phi$, we can get:

$$\begin{aligned} L_\rho(x, z_r, \nu_r, z_c, \nu_c) = \frac{1}{2}\|Ax-y\|_2^2 + \lambda\|z_r\|_1 + \frac{\rho}{2}\|D_r x - z_r + \mu_r\|_2^2 - \frac{\rho}{2}\|\mu_r\|_2^2 \\ + \lambda\|z_c\|_1 + \frac{\rho}{2}\|D_c x - z_c + \mu_c\|_2^2 - \frac{\rho}{2}\|\mu_c\|_2^2 \\ + \phi g(v) + \frac{\Phi}{2}\|x - v + \mu\|_2^2 - \frac{\Phi}{2}\|\mu\|_2^2 \end{aligned} \tag{31}$$

We can update these variables via the following equations:

$$\begin{aligned} x^{(k+1)} = \arg\min_x \ \frac{1}{2}\|Ax^{(k)}-y\|_2^2 + \frac{\rho}{2}\left\|D_r x^{(k)} - z_r^{(k)} + \mu_r^{(k)}\right\|_2^2 + \frac{\rho}{2}\left\|D_c x^{(k)} - z_c^{(k)} + \mu_c^{(k)}\right\|_2^2 + \\ \frac{\Phi}{2}\|x^{(k)} - v^{(k)} + \mu^{(k)}\|_2^2 \end{aligned} \tag{32}$$

$$\begin{aligned} z_r^{(k+1)} &= \arg\min_z \left(\lambda\|z_r^{(k)}\|_1 + \frac{\rho}{2}\left\|D_r x^{(k+1)} - z_r^{(k)} + \mu_r^{(k)}\right\|_2^2\right) \\ z_c^{(k+1)} &= \arg\min_z \left(\lambda\|z_c^{(k)}\|_1 + \frac{\rho}{2}\left\|D_c x^{(k+1)} - z_c^{(k)} + \mu_c^{(k)}\right\|_2^2\right) \\ v^{(k+1)} &= \arg\min_v \left(\phi\ g(v^{(k)}) + \frac{\Phi}{2}\left\|x^{(k)} - v^{(k)} + \mu^{(k)}\right\|_2^2\right) \\ \nu_r^{(k+1)} &= \nu_r^{(k)} + D_r x^{(k+1)} - z_r^{(k+1)} \\ \nu_c^{(k+1)} &= \nu_c^{(k)} + D_c x^{(k+1)} - z_c^{(k+1)} \\ u^{(k+1)} &= u^{(k)} + x^{(k+1)} - v^{(k+1)} \end{aligned} \tag{33}$$

we can regard v subproblem as a denoising problem:

$$\boldsymbol{v}^{(k+1)} = \underset{\boldsymbol{v}\in\mathbb{R}^n}{\operatorname{argmin}} \quad g(\boldsymbol{v}) + \frac{1}{2\sigma^2}\left\|\boldsymbol{v} - \tilde{\boldsymbol{v}}^{(k)}\right\|^2 \tag{34}$$

where $\tilde{v} = x + \mu$.

3.1 x subproblem

Rewrite x subproblem as :

$$x^{(k+1)} = \arg\min_x \ \ \|Ax^{(k)} - y\|_2^2 + \left\|\sqrt{\rho}D_r x^{(k)} - \sqrt{\rho}(z_r^{(k)} - \mu_r^{(k)})\right\|_2^2 \\ + \left\|\sqrt{\rho}D_c x^{(k)} - \sqrt{\rho}(z_c^{(k)} - \mu_c^{(k)})\right\|_2^2 \\ + \|\sqrt{\Phi}x - \sqrt{\Phi}(v^{(k)} - \mu^{(k)})\|_2^2 \tag{35}$$

Write in matrix form :

$$\min_x \left\| \begin{bmatrix} A \\ \sqrt{\rho}D_r \\ \sqrt{\rho}D_c \\ \sqrt{\Phi} \end{bmatrix} x^{(k)} - \begin{bmatrix} y \\ \sqrt{\rho}\left(z_r^{(k)} - \mu_r^{(k)}\right) \\ \sqrt{\rho}\left(z_c^{(k)} - \mu_c^{(k)}\right) \\ \sqrt{\Phi}(v^{(k)} - \mu^{(k)}) \end{bmatrix} \right\|_2^2 \tag{36}$$

use the solution of least square problem $(X^TX)^{-1}X^TY$, we can get:

$$\begin{aligned} x^{(k+1)} &= \left(A^TA + \rho(D_r^TD_r + D_c^TD_c) + \Phi\right)^{-1} \left[A^T, \sqrt{\rho}D_r^T, \sqrt{\rho}D_c^T, \sqrt{\Phi}\right] \begin{bmatrix} y \\ \sqrt{\rho}\left(z_r^{(k)} - \mu_r^{(k)}\right) \\ \sqrt{\rho}\left(z_c^{(k)} - \mu_c^{(k)}\right) \\ \sqrt{\Phi}\left(v^{(k)} - \mu^{(k)}\right) \end{bmatrix} \\ &= \left(A^TA + \rho(D_r^TD_r + D_c^TD_c) + \Phi\right)^{-1} \\ &\quad \left(A^Ty + \rho\left[D_r^T\left(z_r^{(k)} - \mu_r^{(k)}\right) + D_c^T\left(z_c^{(k)} - \mu_c^{(k)}\right)\right] + \Phi\left(v^{(k)} - \mu^{(k)}\right)\right) \end{aligned} \tag{37}$$

4 Deep prior with 3D TV

Objective:

$$\text{minimize} \;\; \frac{1}{2}\|Ax-y\|_2^2 + \phi g(x) + \lambda \sum_i^3 \|D_i x\|_1 \tag{38}$$

Variable substitution:

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}\|Ax-y\|_2^2 + \phi g(v) + \lambda \textstyle\sum_i^3 \|z_i\|_1 \\ \text{subject to} \quad & D_i x - z_i = 0 \\ \text{subject to} \quad & x - v = 0 \end{aligned} \tag{39}$$

Augmented Lagrangian:

$$\begin{aligned} L_\rho(x, z_r, \nu_r, z_c, \nu_c, v) = \frac{1}{2}\|Ax-y\|_2^2 + \sum_i^3 \left(\lambda\|z_i\|_1 + \nu_i^T(D_i x - z_i) + \frac{\rho}{2}\|D_i x - z_i\|_2^2\right) \\ + \phi g(v) + u^T(x-v) + \frac{\beta}{2}\|x-v\|_2^2 \end{aligned} \tag{40}$$

Let $\mu_i = \nu_i/\rho$, $\mu = u/\Phi$, we can get:

$$\begin{aligned} L_\rho(x, z_r, \nu_r, z_c, \nu_c) = \frac{1}{2}\|Ax-y\|_2^2 + \sum_i^3 \left(\lambda\|z_i\|_1 + \frac{\rho}{2}\|D_i x - z_i + \mu_i\|_2^2 - \frac{\rho}{2}\|\mu_i\|_2^2\right) \\ + \phi g(v) + \frac{\beta}{2}\|x - v + \mu\|_2^2 - \frac{\beta}{2}\|\mu\|_2^2 \end{aligned} \tag{41}$$

Optimization:

$$\begin{aligned} x^{(k+1)} &= \arg\min_x \;\; \frac{1}{2}\|Ax^{(k)}-y\|_2^2 + \sum_i^3 \frac{\rho}{2}\left\|D_i x^{(k)} - z_i^{(k)} + \mu_i^{(k)}\right\|_2^2 + \frac{\beta}{2}\|x^{(k)} - v^{(k)} + \mu^{(k)}\|_2^2 \\ z_i^{(k+1)} &= \arg\min_z \left(\lambda\|z_i^{(k)}\|_1 + \frac{\rho}{2}\left\|D_i x^{(k+1)} - z_i^{(k)} + \mu_i^{(k)}\right\|_2^2\right) \\ v^{(k+1)} &= \arg\min_v \left(\phi\; g(v^{(k)}) + \frac{\beta}{2}\left\|x^{(k)} - v^{(k)} + \mu^{(k)}\right\|_2^2\right) \\ \nu_i^{(k+1)} &= \nu_i^{(k)} + D_i x^{(k+1)} - z_i^{(k+1)} \\ u^{(k+1)} &= u^{(k)} + x^{(k+1)} - v^{(k+1)} \end{aligned} \tag{42}$$

4.1 x subproblem

Rewrite x subproblem as :

$$\begin{aligned} x^{(k+1)} = \arg\min_x \;\; \|Ax^{(k)}-y\|_2^2 + \sum_i^3 \left\|\sqrt{\rho} D_i x^{(k)} - \sqrt{\rho}(z_i^{(k)} - \mu_i^{(k)})\right\|_2^2 \\ + \|\sqrt{\beta}x - \sqrt{\beta}(v^{(k)} - \mu^{(k)})\|_2^2 \end{aligned} \tag{43}$$

Write in matrix form :

$$\min_x \left\| \begin{bmatrix} A \\ \sqrt{\rho}D_1 \\ \sqrt{\rho}D_2 \\ \sqrt{\rho}D_3 \\ \sqrt{\beta} \end{bmatrix} x^{(k)} - \begin{bmatrix} y \\ \sqrt{\rho}\left(z_1^{(k)} - \mu_1^{(k)}\right) \\ \sqrt{\rho}\left(z_2^{(k)} - \mu_2^{(k)}\right) \\ \sqrt{\rho}\left(z_3^{(k)} - \mu_3^{(k)}\right) \\ \sqrt{\beta}(v^{(k)} - \mu^{(k)}) \end{bmatrix} \right\|_2^2 \tag{44}$$

use the solution of least square problem $(X^TX)^{-1}X^TY$, we can get:

$$\begin{aligned} x^{(k+1)} &= \left(A^TA + \rho\sum_i^3 D_i^TD_i + \beta\right)^{-1} \left[A^T, \sqrt{\rho}D_1^T, \sqrt{\rho}D_2^T, \sqrt{\rho}D_3^T, \sqrt{\Phi}\right] \begin{bmatrix} y \\ \sqrt{\rho}\left(z_1^{(k)} - \mu_1^{(k)}\right) \\ \sqrt{\rho}\left(z_2^{(k)} - \mu_2^{(k)}\right) \\ \sqrt{\rho}\left(z_3^{(k)} - \mu_3^{(k)}\right) \\ \sqrt{\beta}\left(v^{(k)} - \mu^{(k)}\right) \end{bmatrix} \\ &= \left(A^TA + \rho\sum_i^3 D_i^TD_i + \beta\right)^{-1} \left(A^Ty + \rho\sum_i^3 \left[D_i^T\left(z_i^{(k)} - \mu_i^{(k)}\right)\right] + \beta\left(v^{(k)} - \mu^{(k)}\right)\right) \end{aligned} \tag{45}$$

4.2 v subproblem

$$v^{(k+1)} = \arg\min_v \left(\phi\, g(v^{(k)}) + \frac{\beta}{2}\left\|x^{(k)} - v^{(k)} + \mu^{(k)}\right\|_2^2\right) \tag{46}$$

we can regard v subproblem as a denoising problem:

$$\boldsymbol{v}^{(k+1)} = \underset{\boldsymbol{v}\in\mathbb{R}^n}{\operatorname{argmin}} \quad g(\boldsymbol{v}) + \frac{1}{2\sigma^2}\left\|\boldsymbol{v} - \tilde{\boldsymbol{v}}^{(k)}\right\|^2 \tag{47}$$

where $\tilde{v} = x + \mu$, $\sigma = \sqrt{\phi/\beta}$.

4.3 z subproblem

$$z_i^{(k+1)} = \arg\min_z \left(\lambda\|z_i^{(k)}\|_1 + \frac{\rho}{2}\left\|D_ix^{(k+1)} - z_i^{(k)} + \mu_i^{(k)}\right\|_2^2\right) \tag{48}$$

Since z and v are vectors and v is fixed, this problem is separable and we can solve each 1 dimensional problem individually.

$$\begin{aligned} &\underset{z}{\text{minimize}}\ \lambda\sum_{n=1}^N |z[n]| + \frac{\rho}{2}\sum_{n=1}^N (z[n] - v[n])^2 \\ &= \text{minimize}_z\ \sum_{n=1}^N \left(\lambda|z[n]| + \frac{\rho}{2}(z[n] - v[n])^2\right) \end{aligned} \tag{49}$$

For fixed $v \in R$, we can compute the minimizer of

$$\underset{z\in R}{\text{minimize}}\ \lambda|z| + \frac{\rho}{2}(z - v)^2 \tag{50}$$

explicitly. This function is convex, and is dierentiable everywhere except at z = 0. Away from zero, the derivative is

$$\frac{df}{dz} = \begin{cases} \lambda + z - \rho v, & z > 0 \\ -\lambda + z - \rho v, & z < 0 \end{cases} \tag{51}$$

For the optimal value $z^\star$ to be positive, we need $\lambda + z^\star - \rho v = 0$; this can only hold for $z^\star > 0$ if $v > \lambda/\rho$. Similarly, for $z^\star$ to be negative, we need $-\lambda + z^\star - \rho v = 0$; this can only hold for $z^\star < 0$ if $v < -\lambda/\rho$. If neither of these conditions hold, we must have $z^\star = 0$. Thus

$$z^\star = \begin{cases} \rho v - \lambda, & v > \lambda/\rho \\ 0, & |v| \leq \lambda/\rho \\ \rho v + \lambda, & v < -\lambda/\rho \end{cases} \tag{52}$$

Use $T_{\lambda/\rho}(\cdot)$ to represents this function, we get:

$$z^{(k+1)} = T_{\lambda/\rho}(v) = T_{\lambda/\rho}(Dx + \mu) \tag{53}$$

$T_{\lambda rho}(\cdot)$ is called a **soft thresholding** or **shrinkage** operator

5 Denosing with 3D TV

Objective:

$$\text{minimize} \;\; \frac{1}{2}\|Ax-y\|_2^2 + \lambda \sum_i^3 \|D_i x\|_1 \tag{54}$$

Variable substitution:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax-y\|_2^2 + \lambda \sum_i^3 \|z_i\|_1 \\ \text{subject to} & D_i x - z_i = 0 \end{array} \tag{55}$$

Augmented Lagrangian:

$$L_\rho(x, z_r, \nu_r, z_c, \nu_c, v) = \frac{1}{2}\|Ax-y\|_2^2 + \sum_i^3 \left(\lambda\|z_i\|_1 + \nu_i^T(D_i x - z_i) + \frac{\rho}{2}\|D_i x - z_i\|_2^2\right) \tag{56}$$

Let $\mu_i = \nu_i/\rho$, we can get:

$$L_\rho(x, z_r, \nu_r, z_c, \nu_c) = \frac{1}{2}\|Ax-y\|_2^2 + \sum_i^3 \left(\lambda\|z_i\|_1 + \frac{\rho}{2}\|D_i x - z_i + \mu_i\|_2^2 - \frac{\rho}{2}\|\mu_i\|_2^2\right) \tag{57}$$

Optimization:

$$\begin{aligned} x^{(k+1)} &= \arg\min_x \;\; \frac{1}{2}\|Ax^{(k)}-y\|_2^2 + \sum_i^3 \frac{\rho}{2}\left\|D_i x^{(k)} - z_i^{(k)} + \mu_i^{(k)}\right\|_2^2 \\ z_i^{(k+1)} &= \arg\min_z \left(\lambda\|z_i^{(k)}\|_1 + \frac{\rho}{2}\left\|D_i x^{(k+1)} - z_i^{(k)} + \mu_i^{(k)}\right\|_2^2\right) \\ \nu_i^{(k+1)} &= \nu_i^{(k)} + D_i x^{(k+1)} - z_i^{(k+1)} \end{aligned} \tag{58}$$

5.1 x subproblem

Rewrite x subproblem as :

$$x^{(k+1)} = \arg\min_x \;\; \|Ax^{(k)}-y\|_2^2 + \sum_i^3 \left\|\sqrt{\rho}D_i x^{(k)} - \sqrt{\rho}(z_i^{(k)} - \mu_i^{(k)})\right\|_2^2 \tag{59}$$

Write in matrix form :

$$\min_x \left\| \begin{bmatrix} A \\ \sqrt{\rho}D_1 \\ \sqrt{\rho}D_2 \\ \sqrt{\rho}D_3 \end{bmatrix} x^{(k)} - \begin{bmatrix} y \\ \sqrt{\rho}\left(z_1^{(k)} - \mu_1^{(k)}\right) \\ \sqrt{\rho}\left(z_2^{(k)} - \mu_2^{(k)}\right) \\ \sqrt{\rho}\left(z_3^{(k)} - \mu_3^{(k)}\right) \end{bmatrix} \right\|_2^2 \tag{60}$$

use the solution of least square problem $(X^TX)^{-1}X^TY$, we can get:

$$
\begin{aligned}
x^{(k+1)} &= \left(A^TA + \rho\sum_i^3 D_i^T D_i\right)^{-1} \left[A^T, \sqrt{\rho}D_1^T, \sqrt{\rho}D_2^T, \sqrt{\rho}D_3^T\right] \begin{bmatrix} y \\ \sqrt{\rho}\left(z_1^{(k)} - \mu_1^{(k)}\right) \\ \sqrt{\rho}\left(z_2^{(k)} - \mu_2^{(k)}\right) \\ \sqrt{\rho}\left(z_3^{(k)} - \mu_3^{(k)}\right) \end{bmatrix} \\
&= \left(A^TA + \rho\sum_i^3 D_i^T D_i\right)^{-1} \left(A^Ty + \rho\sum_i^3 \left[D_i^T\left(z_i^{(k)} - \mu_i^{(k)}\right)\right]\right)
\end{aligned} \tag{61}
$$

5.2 z subproblem

$$
z_i^{(k+1)} = \arg\min_z \left(\lambda\|z_i^{(k)}\|_1 + \frac{\rho}{2}\left\|D_i x^{(k+1)} - z_i^{(k)} + \mu_i^{(k)}\right\|_2^2\right) \tag{62}
$$

Since z and v are vectors and v is fixed, this problem is separable and we can solve each 1 dimensional problem individually.

$$
\begin{aligned}
&\underset{z}{\text{minimize}}\ \ \lambda\sum_{n=1}^N |z[n]| + \frac{\rho}{2}\sum_{n=1}^N (z[n] - v[n])^2 \\
=&\,\text{minimize}_z \ \ \sum_{n=1}^N \left(\lambda|z[n]| + \frac{\rho}{2}(z[n] - v[n])^2\right)
\end{aligned} \tag{63}
$$

For fixed $v \in R$, we can compute the minimizer of

$$
\underset{z \in R}{\text{minimize}}\ \ \lambda|z| + \frac{\rho}{2}(z - v)^2 \tag{64}
$$

explicitly. This function is convex, and is differentiable everywhere except at z = 0. Away from zero, the derivative is

$$
\frac{df}{dz} = \begin{cases} \lambda + z - \rho v, & z > 0 \\ -\lambda + z - \rho v, & z < 0 \end{cases} \tag{65}
$$

For the optimal value $z^\star$ to be positive, we need $\lambda + z^\star - \rho v = 0$; this can only hold for $z^\star > 0$ if $v > \lambda/\rho$. Similarly, for $z^\star$ to be negative, we need $-\lambda + z^\star - \rho v = 0$; this can only hold for $z^\star < 0$ if $v < -\lambda/\rho$. If neither of these conditions hold, we must have $z^\star = 0$. Thus

$$
z^\star = \begin{cases} \rho v - \lambda, & v > \lambda/\rho \\ 0, & |v| \le \lambda/\rho \\ \rho v + \lambda, & v < -\lambda/\rho \end{cases} \tag{66}
$$

Use $T_{\lambda/\rho}(\cdot)$ to represents this function, we get:

$$
z^{(k+1)} = T_{\lambda/\rho}(v) = T_{\lambda/\rho}(Dx + \mu) \tag{67}
$$

$T_{\lambda rho}(\cdot)$ is called a **soft thresholding** or **shrinkage** operator

6 Enhanced 3D TV regularized DPHSIR

Objective:

$$
\begin{array}{rl}
\text{minimize} & \lambda \sum_i^3 \|U_i\|_1 + \|E\|_1 + \phi g(T) \\
\text{subject to} & Y = AX + E \\
\text{subject to} & D_i X = U_i V_i^T, V_i^T V_i = I \\
\text{subject to} & T = X
\end{array} \tag{68}
$$

Augmented Lagrangian

$$
\begin{aligned}
L() =& \lambda \sum_i^3 \|U_i\|_1 + \|E\|_1 + \phi g(T) + \\
&+ \sum_i^3 \Big(\lambda \|U_i\|_1 + M_i^T (D_i X - U_i V_i^T) + \frac{\mu}{2} \|D_i X - U_i V_i^T\|_2^2 \Big) \\
&+ \Gamma^T (Y - AX - E) + \frac{\mu}{2} \|Y - AX - E\|_2^2 \\
&+ \phi g(T) + Q^T (X - T) + \frac{\mu}{2} \|X - T\|_2^2
\end{aligned} \tag{69}
$$