# 毕设优化算法初步设计及相应问题

> 汇报人：周雄
>
> 日期：2018-01-20

## Definition

**Main goal**: *Retrievaling phase from some intensity observations with their own distance.*

**others**: *Obtaining better quality and robustness by denoising with a TV prior*

$$\mathbf{o}_r = |\mathbf{A}_r \cdot \mathbf{u}_0|^2 + \varepsilon_r, r = 1, 2, \ldots, K$$

## The optimization problem

$$\hat{\mathbf{u}}_0 = \arg\min_{\mathbf{u}_0} \sum_{r=1}^{K} \frac{1}{2\sigma_r^2} \|\mathbf{o}_r - |\mathbf{A}_r \mathbf{u}_0|^2\|_2^2 + \mu \cdot \|\mathbf{u}_0\|_{TV}$$

where $\mathbf{o}_r \in \mathbb{R}^{m\times n}, \mathbf{u}_r \in \mathbb{C}^{m\times n}, \mathbf{A}_r \in \mathbb{C}^{m\times m}$, $|\cdot|$ is the module, $\|\cdot\|$ is the norm, and $\|\cdot\|_{TV}$ represents the TV penality.

## The TV penality

Given $\mathbf{u} \in \mathbb{C}^{m \times n}$, the gradient $\nabla \mathbf{u}$ (including vertical and horizontal gradients) is a vector in $\mathbb{C}^{m \times n} \times \mathbb{C}^{m \times n}$ calculated by

$$(\nabla \mathbf{u})_{i,j} = \left((\nabla \mathbf{u})_{i,j}^{v}, (\nabla \mathbf{u})_{i,j}^{h}\right)$$

with

$$(\nabla \mathbf{u})_{i,j}^{v} = \begin{cases} \mathbf{u}_{i+1,j} - \mathbf{u}_{i,j} & \text{if } i < m \\ 0 & \text{if } i = m \end{cases}$$

$$(\nabla \mathbf{u})_{i,j}^{h} = \begin{cases} \mathbf{u}_{i,j+1} - \mathbf{u}_{i,j} & \text{if } j < n \\ 0 & \text{if } j = n \end{cases}$$

Then the Total Varaition is

$$\|\mathbf{u}_0\|_{TV} = \sum_{i,j} \|(\nabla \mathbf{u}_0)_{i,j}\|$$

where the norm $\|\cdot\|$ can be the L1-norm, L2-norm or others. The TV is **isotropic** if the norm in the summation is L2-norm and **anisotropic** when L1-norm. **Here we consider the L1-norm firstly**, and the optimization problem can be formed as

$$\hat{\mathbf{u}}_0 = \arg\min_{\mathbf{u}_0} \sum_{r=1}^{K} \frac{1}{2\sigma_r^2} \|\mathbf{o}_r - |\mathbf{A}_r \mathbf{u}_0|^2\|_2^2 + \mu \sum_{i,j} \|(\nabla \mathbf{u}_0)_{i,j}\|_1$$

For simplicity, we mark $(\nabla \mathbf{u})^v,\ (\nabla \mathbf{u})^h$ as $\mathbf{D}_v \mathbf{u}, \mathbf{D}_h \mathbf{u}$, respectively.

$$\hat{\mathbf{u}}_0 = \arg\min_{\mathbf{u}_0} \sum_{r=1}^{K} \frac{1}{2\sigma_r^2} \|\mathbf{o}_r - |\mathbf{A}_r \mathbf{u}_0|^2\|_2^2 + \mu \|\mathbf{D}_v \mathbf{u}_0\|_1 + \mu \|\mathbf{D}_h \mathbf{u}_0\|_1$$

# The TV is L1-norm

## Adding variables-splitting techinique to the optimization problem

$$\hat{\mathbf{u}}_0 = \arg\min_{\mathbf{u}_0} \sum_{r=1}^{K} \frac{1}{2\sigma_r^2} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|_2^2 + \mu \|\mathbf{u}_v\|_1 + \mu \|\mathbf{u}_h\|_1 \ \text{ s.t. } \ \mathbf{u}_v = \mathbf{D}_v \mathbf{u}_0,\ \mathbf{u}_h = \mathbf{D}_h \mathbf{u}_0,\ \mathbf{u}_r = \mathbf{A}_r \cdot \mathbf{u}_0, r = 1, 2, \ldots, K$$

## Using the augmented Lagrangian (AL) method

$$
\mathcal{L} = \sum_{r=1}^{K} \frac{1}{2\sigma_r^2}\left[\left\|\mathbf{o}_r - |\mathbf{u}_r|^2\right\|_2^2 + \frac{\lambda_r}{2}\|\mathbf{A}_r\mathbf{u}_0 - \mathbf{u}_r - \mathbf{v}_r\|_2^2 - \frac{\lambda_r}{2}\|\mathbf{v}_r\|_2^2\right] + \mu\|\mathbf{u}_v\|_1 + \frac{\lambda_v}{2}\|\mathbf{D}_v\mathbf{u}_0 - \mathbf{u}_v - \mathbf{v}_v\|_2^2 - \frac{\lambda_v}{2}\|\mathbf{v}_v\|_2^2 + \mu\|\mathbf{u}_h\|_1 + \frac{\lambda_h}{2}\|\mathbf{D}_h\mathbf{u}_0 - \mathbf{u}_h - \mathbf{v}_h\|_2^2 - \frac{\lambda_h}{2}\|\mathbf{v}_h\|_2^2
$$

The augumented Lagrangian form is called the scaled-version of ADMM's.

## The iterating steps

$$
\begin{aligned}
\mathbf{u}_0^{t+1} &= \arg\min_{\mathbf{u}_0} \mathcal{L}(\mathbf{u}_0, \{\mathbf{u}_r^t\}, \{\mathbf{v}_r^t\}, \mathbf{u}_v^t, \mathbf{v}_v^t, \mathbf{u}_h^t, \mathbf{v}_h^t) = \arg\min_{\mathbf{u}_0} \sum_{r=1}^{K} \frac{\lambda_r}{4\sigma_r^2}\|\mathbf{A}_r\mathbf{u}_0 - \mathbf{u}_r^t - \mathbf{v}_r^t\|_2^2 + \frac{\lambda_v}{2}\|\mathbf{D}_v\mathbf{u}_0 - \mathbf{u}_v - \mathbf{v}_v\|_2^2 \\
&\quad + \frac{\lambda_h}{2}\|\mathbf{D}_h\mathbf{u}_0 - \mathbf{u}_h - \mathbf{v}_h\|_2^2 \\
\mathbf{u}_r^{t+1} &= \arg\min_{\mathbf{u}_r} \mathcal{L}(\mathbf{u}_0^{t+1}, \{\mathbf{u}_r\}, \{\mathbf{v}_r^t\}) = \arg\min_{\mathbf{u}_r} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|_2^2 + \frac{\lambda_r}{2}\|\mathbf{A}_r\mathbf{u}_0^{t+1} - \mathbf{u}_r - \mathbf{v}_r^t\|_2^2 \\
\mathbf{u}_v^{t+1} &= \arg\min_{\mathbf{u}} \mathcal{L}(\mathbf{u}_0^{t+1}, \mathbf{u}_v^t, \mathbf{v}_v^t) = \arg\min_{\mathbf{u}_v} \mu\|\mathbf{u}_v\|_1 + \frac{\lambda_v}{2}\|\mathbf{D}_v\mathbf{u}_0^{t+1} - \mathbf{u}_v - \mathbf{v}_v^t\|_2^2 \\
\mathbf{u}_h^{t+1} &= \arg\min_{\mathbf{u}} \mathcal{L}(\mathbf{u}_0^{t+1}, \mathbf{u}_h^t, \mathbf{v}_h^t) = \arg\min_{\mathbf{u}_h} \mu\|\mathbf{u}_h\|_1 + \frac{\lambda_v}{2}\|\mathbf{D}_h\mathbf{u}_0^{t+1} - \mathbf{u}_h - \mathbf{v}_h^t\|_2^2 \\
\mathbf{v}_r^{t+1} &= \mathbf{v}_r^t - (\mathbf{A}_r\mathbf{u}_0^{t+1} - \mathbf{u}_r^{t+1}) \\
\mathbf{v}_v^{t+1} &= \mathbf{v}_v^t - (\mathbf{D}_v\mathbf{u}_0^{t+1} - \mathbf{u}_v^{t+1}) \\
\mathbf{v}_h^{t+1} &= \mathbf{v}_h^t - (\mathbf{D}_h\mathbf{u}_0^{t+1} - \mathbf{u}_h^{t+1})
\end{aligned}
$$

## For $\mathbf{u}_0$

$$
\begin{aligned}
0 = \frac{\partial \mathcal{L}(\mathbf{u}_0, \{\mathbf{u}_r^t\}, \{\mathbf{v}_r^t\}, \mathbf{u}^t, \mathbf{v}^t)}{\partial \mathbf{u}_0} &= \sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H (\mathbf{A}_r \mathbf{u}_0 - \mathbf{u}_r^t - \mathbf{v}_r^t) + \lambda_v \mathbf{D}_v^H (\mathbf{D}_v \mathbf{u}_0 - \mathbf{u}_v^t - \mathbf{v}_v^t) + \lambda_h \mathbf{D}_h^H (\mathbf{D}_h \mathbf{u}_0 - \mathbf{u}_h^t - \mathbf{v}_h^t) \\
\Rightarrow \mathbf{u}_0^{t+1} &= \left( \sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H \mathbf{A}_r + \lambda_v \mathbf{D}_v^H \mathbf{D}_v + \lambda_h \mathbf{D}_h^H \mathbf{D}_h \right)^{-1} \left( \sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H (\mathbf{u}_r^t + \mathbf{v}_r^t) + \lambda_v \mathbf{D}_v (\mathbf{u}_v^t + \mathbf{v}_v^t) + \lambda_h \mathbf{D}_h (\mathbf{u}_h^t + \mathbf{v}_h^t) \right)
\end{aligned}
$$

where $\mathbf{A}_r^H$ is the Hermitian transpose of $\mathbf{A}_r$.

## For $\mathbf{u}_r$

The objective function is additive w.r.t. the matrixs $\mathbf{u}_r$ and their components. Thus, the minimization on $\mathbf{u}_r$ can be produced in the elementwise manner. The derivative $\nabla_{\mathbf{u}_r^*[i,j]} \mathcal{L} = 0$ gives the minimum condition for $\mathbf{u}_r^*[i,j]$ as

$$
\begin{aligned}
\frac{\partial \mathcal{L}(\mathbf{u}_0^{t+1}, \{\mathbf{u}_r\}, \{\mathbf{v}_r^t\})}{\partial \mathbf{u}_r^*[i,j]} &= 4 \left( |\mathbf{u}_r[i,j]|^2 - \mathbf{o}_r[i,j] \right) \cdot \mathbf{u}_r[i,j] + \lambda_r (\mathbf{u}_r[i,j] - \mathbf{A}_r \mathbf{u}_0^{t+1}[i,j] + \mathbf{v}_r^t[i,j]) = 0 \\
\Rightarrow \mathbf{u}_r[i,j] &= \frac{\mathbf{A}_r \mathbf{u}_0^{t+1}[i,j] - \mathbf{v}_r^t[i,j]}{\frac{4}{\lambda_r} \left( |\mathbf{u}_r[i,j]|^2 - \mathbf{o}_r[i,j] \right) + 1} = \frac{\eta_r[i,j]}{\kappa_r[i,j]} (\text{Molduling on both sides}) \\
\Rightarrow |\mathbf{u}_r[i,j]|^3 &+ |\mathbf{u}_r[i,j]| \cdot \left( \frac{\lambda_r}{4} - \mathbf{o}_r[i,j] \right) - \frac{\lambda_r}{4} \operatorname{sgn}(\kappa_r[i,j]) \cdot |\eta_r[i,j]| = 0
\end{aligned}
$$

In reality, we have two different cubic equations: corresponding to $\operatorname{sgn}(\kappa_r[i,j]) = 1$ and to $\operatorname{sgn}(\kappa_r[i,j]) = -1$. Each of these may have a single or three real solutions. We are looking for a nonnegative real root denotes as $\widetilde{\mathbf{u}}[i,j]|$. It can be seen that such $\widetilde{\mathbf{u}}[i,j]$ always exists.

After finding the root, the corresponding complex-valued estimate of the wave field at the sensor plane $\hat{\mathbf{u}}_r[i,j]$ is calculated as

$$\hat{\mathbf{u}}_r[i,j] = \frac{\mathbf{A}_r\mathbf{u}_0^{t+1}[i,j] - \mathbf{v}_r^t[i,j]}{\frac{4}{\lambda_r}(|\widetilde{\mathbf{u}}_r[i,j]|^2 - \mathbf{o}_r[i,j]) + 1}$$

Then we can update $\mathbf{u}_r$ in

$$\hat{\mathbf{u}}_r = (\mathbf{A}_r\mathbf{u}_0^{t+1} - \mathbf{v}_r^t) \oslash (\frac{4}{\lambda}(|\widetilde{\mathbf{u}}_r|^2 - \mathbf{o}_r) + 1)$$

where operator $\oslash$ is the element-by-element quotient of two matrix.

# For $\mathbf{u}_v, \mathbf{u}_h$

Updating $\mathbf{u}_v,\ \mathbf{u}_h$ by solving the sub-problem:

$$\mathbf{u}^{t+1} = \arg\min_{\mathbf{u}} \mu\|\mathbf{u}\|_1 + \frac{\lambda}{2}\|\mathbf{u} - \mathbf{s}\|_2^2$$

What's more, to simplify the question, we decide to some marks below:

$$\mathbf{u} = \mathbf{u}_a + i \cdot \mathbf{u}_b, \mathbf{s} = \mathbf{s}_a + i \cdot \mathbf{s}_b$$

So the above optimization of $u$ can be rewritten as

$$\mathbf{u} = \arg\min_{\mathbf{u}_a,\mathbf{u}_b} \mu\|\mathbf{u}_a + i \cdot \mathbf{u}_b\|_1 + \frac{\lambda}{2}\|\mathbf{u}_a - \mathbf{s}_a\|_2^2 + \frac{\lambda}{2}\|\mathbf{u}_b - \mathbf{s}_b\|_2^2$$

Then we can get the solution by substituting the corresponding variables and parameters into the above method.

## The sub-problem's optimization method

1) Suppose that $\|\mathbf{u}\|_1 = \|\mathbf{u}_a\|_1 + \|\mathbf{u}_b\|_1$

then we need to solve two sub-questions:

$$\hat{\mathbf{u}}_a = \arg\min_{\mathbf{u}_a} \mu\|\mathbf{u}_a\|_1 + \frac{\lambda}{2}\|\mathbf{u}_a - \mathbf{s}_a\|_2^2$$
$$\hat{\mathbf{u}}_b = \arg\min_{\mathbf{u}_b} \mu\|\mathbf{u}_b\|_1 + \frac{\lambda}{2}\|\mathbf{u}_b - \mathbf{s}_b\|_2^2$$

We can use the ST(Soft Threholding) method get

$$\hat{\mathbf{u}}_a = \mathrm{soft}(\mathbf{s}_a, \frac{2\mu}{\lambda}) = \mathrm{sgn}(\mathbf{s}_a) \odot \max\left\{|\mathbf{s}_a| - \frac{\mu}{\lambda}, 0\right\}$$
$$\hat{\mathbf{u}}_b = \mathrm{soft}(\mathbf{s}_b, \frac{2\mu}{\lambda}) = \mathrm{sgn}(\mathbf{s}_b) \odot \max\left\{|\mathbf{s}_b| - \frac{\mu}{\lambda}, 0\right\}$$

In [2]:
```
def soft(s, mu, gamma, sep=True):
        sa, sb = np.real(s), np.imag(s)
        if sep:
                ua = np.sign(sa) * np.maximum(np.abs(sa) - mu / gamma, 0)
                ub = np.sign(sb) * np.maximum(np.abs(sb) - mu / gamma, 0)
        else:
                coef = np.sqrt(sa**2 / (sa**2+sb**2))
                ua = coef * np.sign(sa) * np.maximum(np.sqrt(sa**2 + sb **2) - mu / gamma, 0)
                ub = sb / sa * ua

        return ua + ub * 1j
```

(2) Suppose that $\|\mathbf{u}\|_1 = \left\|\sqrt{\mathbf{u}_a^2+\mathbf{u}_b^2}\right\|_1$

$$\mathbf{u} = \arg\min_{\mathbf{u}_a,\mathbf{u}_b} \mu\cdot\left\|\sqrt{\mathbf{u}_a^2+\mathbf{u}_b^2}\right\|_1 + \frac{\lambda}{2}\|\mathbf{u}_a-\mathbf{s}_a\|_2^2 + \frac{\lambda}{2}\|\mathbf{u}_b-\mathbf{s}_b\|_2^2 = \arg\min_{\mathbf{u}_a,\mathbf{u}_b}\mathcal{L}(\mathbf{u}_a,\mathbf{u}_b)$$

The derivative $\nabla_{\mathbf{u}_a[i,j]}\mathcal{L}$ and $\nabla_{\mathbf{u}_b[i,j]}\mathcal{L}$ are equal to 0.So we have

$$\frac{\partial\mathcal{L}}{\partial\mathbf{u}_a[i,j]} = \frac{\mu\cdot\mathbf{u}_a[i,j]}{\sqrt{\mathbf{u}_a^2[i,j]+\mathbf{u}_b^2[i,j]}} + \lambda(\mathbf{u}_a[i,j]-\mathbf{s}_a[i,j]) = 0$$
$$\frac{\partial\mathcal{L}}{\partial\mathbf{u}_b[i,j]} = \frac{\mu\cdot\mathbf{u}_b[i,j]}{\sqrt{\mathbf{u}_a^2[i,j]+\mathbf{u}_b^2[i,j]}} + \lambda(\mathbf{u}_b[i,j]-\mathbf{s}_b[i,j]) = 0$$

So we get

$$\frac{\mathbf{u}_b[i,j]}{\mathbf{u}_a[i,j]} = \frac{\mathbf{s}_b[i,j]-\mathbf{u}_b[i,j]}{\mathbf{s}_a[i,j]-\mathbf{u}_a[i,j]} \Rightarrow \frac{\mathbf{u}_b[i,j]}{\mathbf{u}_a[i,j]} = \frac{\mathbf{s}_b[i,j]}{\mathbf{s}_a[i,j]}$$

Then we obtain that

$$\hat{\mathbf{u}}_a[i,j] = \arg\min_{\mathbf{u}_a[i,j]} \mu\cdot\sqrt{\mathbf{1}+\frac{\mathbf{s}_b^2[i,j]}{\mathbf{s}_a^2[i,j]}}\,|\mathbf{u}_a[i,j]| + \frac{\lambda}{2}(\mathbf{u}_a[i,j]-\mathbf{s}_a[i,j])^2 + \frac{\lambda}{2}\left(\frac{\mathbf{s}_b[i,j]}{\mathbf{s}_a[i,j]}\mathbf{u}_a[i,j]-\mathbf{s}_b[i,j]\right)^2$$

marking $a=\mathbf{s}_a[i,j], b=\mathbf{s}_b[i,j], x=\mathbf{u}_a[i,j]$. The above question can be simplied as

$$\hat{\mathbf{u}}_a[i,j] = \arg\min_{\mathbf{x}} \mu\cdot\left|\sqrt{\frac{a^2+b^2}{a^2}}\cdot x\right| + \frac{\lambda}{2}\left(\sqrt{\frac{a^2+b^2}{a^2}}\cdot x - \mathrm{sgn}(a)\sqrt{a^2+b^2}\right)^2$$

We also can use ST method to solve the above question, and we get

$$\sqrt{\frac{a^2+b^2}{a^2}}\hat{x} = \mathrm{sgn}(\mathrm{sgn}(a)\sqrt{a^2+b^2})\cdot\max\left\{|\mathrm{sgn}(a)\sqrt{a^2+b^2}| - \frac{\mu}{\lambda}, 0\right\} = \mathrm{sgn}(a)\cdot\max\left\{\sqrt{a^2+b^2}-\frac{\mu}{\lambda},0\right\}$$

So, we get the update result of $\mathbf{u}_a[i,j], \mathbf{u}_b[i,j]$

$$\hat{\mathbf{u}}_a[i,j] = \sqrt{\frac{\mathbf{s}_a^2[i,j]}{\mathbf{s}_a^2[i,j]+\mathbf{s}_b^2[i,j]}}\mathrm{sgn}(\mathbf{s}_a[i,j])\max\left\{\sqrt{\mathbf{s}_a^2[i,j]+\mathbf{s}_b^2[i,j]}-\frac{\mu}{\lambda},0\right\}$$
$$\hat{\mathbf{u}}_b[i,j] = \frac{\mathbf{s}_b[i,j]}{\mathbf{s}_a[i,j]}\sqrt{\frac{\mathbf{s}_a^2[i,j]}{\mathbf{s}_a^2[i,j]+\mathbf{s}_b^2[i,j]}}\mathrm{sgn}(\mathbf{s}_a[i,j])\max\left\{\sqrt{\mathbf{s}_a^2[i,j]+\mathbf{s}_b^2[i,j]}-\frac{\mu}{\lambda},0\right\}$$

## The experiment

We assume that $\sigma_r^2 = \sigma^2$, $\mathbf{A}_r^H \mathbf{A}_r = \mathbf{I}$, for $\mathbf{u}_0$, we have

$$\mathbf{u}_0^{t+1} = \left( \sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H \mathbf{A}_r + \lambda_v \mathbf{D}_v^H \mathbf{D}_v + \lambda_h \mathbf{D}_h^H \mathbf{D}_h \right)^{-1} \left( \sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H (\mathbf{u}_r^t + \mathbf{v}_r^t) + \lambda_v \mathbf{D}_v (\mathbf{u}_v^t + \mathbf{v}_v^t) + \lambda_h \mathbf{D}_h (\mathbf{u}_h^t + \mathbf{v}_h^t) \right)$$

$$\Rightarrow \mathbf{u}_0^{t+1} = \left( \sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{I}_{m \times m} + \lambda_v \mathbf{D}_v^H \mathbf{D}_v + \lambda_h \mathbf{D}_h^H \mathbf{D}_h \right)^{-1} \left( \sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H (\mathbf{u}_r^t + \mathbf{v}_r^t) + \lambda_v \mathbf{D}_v (\mathbf{u}_v^t + \mathbf{v}_v^t) + \lambda_h \mathbf{D}_h (\mathbf{u}_h^t + \mathbf{v}_h^t) \right)$$

# The TV is L2-norm