毕设优化算法初步设计及相应问题

> 汇报人：周雄
>
> 日期：2019-02-15

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

Based on maximum Likelihood Estimation (MLE) and the concept of penalty term (or just on Maximum a posteriori estimation, MAP), we can get

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

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 $\mathrm{pen}(\cdot)$ represents the penality.

Adding variables-splitting techinique to the optimization problem

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

Using the augmented Lagrangian (AL) method

$$\mathcal{L} = \sum_{r=1}^{K} \frac{1}{2\sigma_r^2} \left[\left\| \mathbf{o}_r - |\mathbf{u}_r|^2 \right\|_2^2 + \frac{\lambda_r}{2} \|\mathbf{A}_r \mathbf{u}_0 - \mathbf{u}_r - \mathbf{v}_r\|_2^2 - \frac{\lambda_r}{2} \|\mathbf{v}_r\|_2^2 \right] + \mu \cdot \mathrm{pen}(\mathbf{u}) + \frac{\lambda}{2} \|\mathbf{u}_0 - \mathbf{u} - \mathbf{v}\|_2^2 - \frac{\lambda}{2} \|\mathbf{v}\|_2^2$$

The augumented Lagrangian form is called the **scaled-version** of ADMM (Boyd et al., 2011). Although the optimization problem is proposed for complex domain, the scaled-version is also applicable (because we can regard the complex-value variable as a two-dimensional variable under the definition of the norm). What's more, the scaled version pays attention to both real part and imaginary part, so the retrieval should be better than the former methods.

The ADMM iterating steps

In this case, the ADMM can be written as

$$\begin{aligned}
\mathbf{u}_0^{t+1} &= \arg\min_{\mathbf{u}_0} \mathcal{L}(\mathbf{u}_0, \{\mathbf{u}_r^t\}, \{\mathbf{v}_r^t\}, \mathbf{u}^t, \mathbf{v}^t) = \arg\min_{\mathbf{u}_0} \sum_{r=1}^{K} \frac{\lambda_r}{4\sigma_r^2} \|\mathbf{A}_r \mathbf{u}_0 - \mathbf{u}_r^t - \mathbf{v}_r^t\|_2^2 + \frac{\lambda}{2} \|\mathbf{u}_0 - \mathbf{u} - \mathbf{v}\|_2^2 \\
\mathbf{u}_r^{t+1} &= \arg\min_{\mathbf{u}_r} \mathcal{L}(\mathbf{u}_0^{t+1}, \{\mathbf{u}_r\}, \{\mathbf{v}_r^t\}) = \arg\min_{\mathbf{u}_r} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|_2^2 + \frac{\lambda_r}{2} \|\mathbf{A}_r \mathbf{u}_0^{t+1} - \mathbf{u}_r - \mathbf{v}_r^t\|_2^2 \\
\mathbf{u}^{t+1} &= \arg\min_{\mathbf{u}} \mathcal{L}(\mathbf{u}_0^{t+1}, \mathbf{u}^t, \mathbf{v}^t) = \arg\min_{\mathbf{u}} \mu \cdot \mathrm{pen}(\mathbf{u}) + \frac{\lambda}{2} \|\mathbf{u}_0^{t+1} - \mathbf{u} - \mathbf{v}^t\|_2^2 \\
\mathbf{v}_r^{t+1} &= \mathbf{v}_r^t - (\mathbf{A}_r \mathbf{u}_0^{t+1} - \mathbf{u}_r^{t+1}) \\
\mathbf{v}^{t+1} &= \mathbf{v}^t - (\mathbf{u}_0^{t+1} - \mathbf{u}^{t+1})
\end{aligned}$$

Details of specific updates

For $\mathbf{u}_0$

$$
\begin{aligned}
0 = \frac{\partial \mathcal{L}(\mathbf{u}_0, \{\mathbf{u}_r^t\}, \{\mathbf{v}_r^t\}, \mathbf{u}^t, \mathbf{v}^t)}{\partial \mathbf{u}_0} &= \sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H (\mathbf{A}_r \mathbf{u}_0 - \mathbf{u}_r^t - \mathbf{v}_r^t) + \lambda(\mathbf{u}_0 - \mathbf{u}^t - \mathbf{v}^t) \\
\Rightarrow \mathbf{u}_0^{t+1} &= \left(\sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H \mathbf{A}_r + \lambda \cdot \mathbf{I}_{m\times m} \right)^{-1} \left(\sum_{r=1}^{K} \frac{\lambda_r}{2\sigma_r^2} \mathbf{A}_r^H (\mathbf{u}_r^t + \mathbf{v}_r^t) + \lambda(\mathbf{u}^t + \mathbf{v}^t) \right)
\end{aligned}
$$

where $\mathbf{A}_r^H$ is **the Hermitian transpose (conjugate transpose)** of $\mathbf{A}_r$. Actually, $\mathbf{A}_r^H \mathbf{A}_r$ is the unit matrix

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

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

In reality, we have two different cubic equations: corresponding to $\mathrm{sgn}(\kappa_r[i,j]) = 1$ and to $\mathrm{sgn}(\kappa_r[i,j]) = -1$. 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, because while $\hat{x}$ is the solution of the equation $x^3 + px + q = 0$, $-\hat{x}$ is the solution of $x^3 + px - q = 0$.

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

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

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

For $\mathbf{u}$

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

$$\mathbf{u}^{t+1} = \arg\min_{\mathbf{u}} \mu \cdot \mathrm{pen}(\mathbf{u}) + \frac{\lambda}{2} \|\mathbf{u}_0^{t+1} - \mathbf{u} - \mathbf{v}^t\|_2^2$$

As for the choice of $\mathrm{pen}(\cdot)$, we pay more attention to **L1-norm** and consider two situations.

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

$$\mathbf{u} = \mathbf{u}_a + i \cdot \mathbf{u}_b,$$
$$\mathbf{s} = \mathbf{u}_0^{t+1} - \mathbf{v}^t = \mathbf{s}_a + i \cdot \mathbf{s}_b$$

So the above optimization of $\mathbf{u}$ can be rewritten as

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

L1-norm: $\mathrm{pen}(\mathbf{u}) = \|\mathbf{u}\|_1$

the above optimization problem can be written as

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

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

then we just 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 and obtain

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

Then we get the updated $\mathbf{u}^{t+1} = \hat{\mathbf{u}}_a + i \cdot \hat{\mathbf{u}}_b$.

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

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

Considering the derivative $\nabla_{\mathbf{u}_a[i,j]}\mathcal{L}$ and $\nabla_{\mathbf{u}_b[i,j]}\mathcal{L}$ as 0So 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

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

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

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

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

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

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

The experiment

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

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

$$\Rightarrow \mathbf{u}_0^{t+1} = \left(\sum_{r=1}^{K} \lambda_r + 2\lambda\sigma^2 \right)^{-1} \left(\sum_{r=1}^{K} \lambda_r \mathbf{A}_r^H (\mathbf{u}_r^t + \mathbf{v}_r^t) + 2\lambda\sigma^2 (\mathbf{u}^t + \mathbf{v}^t) \right)$$

And we simplify some parameters to get more structral and more beautiful calculation

$$\tilde{\lambda}_r = \frac{\lambda_r}{\sum_{r=1}^{K} \lambda_r + 2\lambda\sigma^2}, \tilde{\lambda} = \frac{2\lambda\sigma^2}{\sum_{r=1}^{K} \lambda_r + 2\lambda\sigma^2}, \gamma = \sum_{r=1}^{K} \lambda_r + 2\lambda\sigma^2$$

Then we obtain the specific iterating steps:

$$\begin{aligned}
\mathbf{u}_0^{t+1} &= \sum_{r=1}^{K} \tilde{\lambda}_r \mathbf{A}_r^H (\mathbf{u}_r^t + \mathbf{v}_r^t) + \tilde{\lambda}(\mathbf{u}^t + \mathbf{v}^t) \\
\mathbf{u}_r^{t+1} &= (\mathbf{A}_r \mathbf{u}_0^{t+1} - \mathbf{v}_r^t) \oslash (\frac{4}{\gamma \cdot \tilde{\lambda}_r} (|\tilde{\mathbf{u}}_r|^2 - \mathbf{o}_r) + 1) \\
\mathbf{u}^{t+1} &= \mathrm{soft}(\mathbf{u}_0^{t+1} - \mathbf{v}^t,\ \mu,\ 2\tilde{\lambda}\gamma\sigma^2) \\
\mathbf{v}_r^{t+1} &= \mathbf{v}_r^t - (\mathbf{A}_r \mathbf{u}_0^{t+1} - \mathbf{u}_r^{t+1}) \\
\mathbf{v}^{t+1} &= \mathbf{v}^t - (\mathbf{u}_0^{t+1} - \mathbf{u}^{t+1})
\end{aligned}$$

Actually, we set $K = 5,\ \tilde{\lambda} = 0.1,\ \tilde{\lambda}_r = (1 - \tilde{\lambda})/K,\ \mu = 5e-2,\ \gamma = 2$ by default.