

Consider the optimization problem:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & \frac{1}{2} \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_1 \\ \text{subject to } & \mathbf{z} - \mathbf{x} = 0 \end{aligned} \quad (1)$$

The Lagrangian function has a saddlepoint at the solution. In traditional ADMM, the Lagrangian function

$$\mathbf{L}(\mathbf{x}, \mathbf{z}, \gamma) = \frac{1}{2} \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_1 + \gamma^H (\mathbf{z} - \mathbf{x}) \quad (2)$$

is augmented with a term $\frac{1}{\rho} \|\mathbf{z} - \mathbf{x}\|_2^2$. This added term does not change the saddlepoint solution; for all feasible points, the added term is zero. It's purpose is to make finding the saddlepoint easier. The resulting function is known as the augmented Lagrangian.

$$\mathbf{L}_\rho(\mathbf{x}, \mathbf{z}, \gamma) = \frac{1}{2} \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_1 + \gamma^H (\mathbf{z} - \mathbf{x}) + \frac{1}{\rho} \|\mathbf{z} - \mathbf{x}\|_2^2 \quad (3)$$

Now, instead of augmenting the lagrangian with the traditional term $\frac{\rho}{2} \|\mathbf{z} - \mathbf{x}\|_2^2$, consider an alternative term $\frac{1}{2} \|\mathbf{\Lambda}^{\frac{1}{2}} (\mathbf{z} - \mathbf{x})\|_2^2$ where $\mathbf{\Lambda}$ is a diagonal positive definite matrix.

Then,

$$\mathbf{L}_\mathbf{\Lambda}(\mathbf{x}, \mathbf{z}, \gamma) = \frac{1}{2} \|\mathbf{s} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_1 + \gamma^H (\mathbf{z} - \mathbf{x}) + \frac{1}{2} \|\mathbf{\Lambda}^{\frac{1}{2}} (\mathbf{z} - \mathbf{x})\|_2^2 \quad (4)$$

$$\nabla_{\mathbf{x}} \mathbf{L}_\mathbf{\Lambda}(\mathbf{x}, \mathbf{z}, \gamma) = \mathbf{D}^H \mathbf{D}\mathbf{x} - \mathbf{D}^H \mathbf{s} + \mathbf{\Lambda}\mathbf{x} - \mathbf{\Lambda}\mathbf{z} - \gamma \quad (5)$$

So,

$$\arg \min_{\mathbf{x}} \mathbf{L}_\mathbf{\Lambda}(\mathbf{x}, \mathbf{z}, \gamma) = (\mathbf{\Lambda} + \mathbf{D}^H \mathbf{D})^{-1} (\mathbf{D}^H \mathbf{s} - \mathbf{\Lambda}\mathbf{z} - \gamma) \quad (6)$$

and

$$\arg \min_{\mathbf{z}} \mathbf{L}_\mathbf{\Lambda}(\mathbf{x}, \mathbf{z}, \gamma) = \mathbf{S}_{\lambda\mathbf{\Lambda}^{-1}}(\mathbf{x} - \mathbf{\Lambda}^{-1}\gamma) \quad (7)$$

Therefore, we get the following ADMM iterative update equations:

$$\mathbf{x}^{(k+1)} = (\mathbf{\Lambda} + \mathbf{D}^H \mathbf{D})^{-1} (\mathbf{D}^H \mathbf{s} - \mathbf{\Lambda}\mathbf{z}^{(k)} - \gamma^{(k)}) \quad (8)$$

$$\mathbf{z}^{(k+1)} = \mathbf{S}_{\lambda\mathbf{\Lambda}^{-1}}(\mathbf{x}^{(k+1)} - \mathbf{\Lambda}^{-1}\gamma^{(k)}) \quad (9)$$

$$\gamma^{(k+1)} = \gamma^{(k)} + \mathbf{\Lambda}(\mathbf{z}^{(k+1)} - \mathbf{x}^{(k+1)}) \quad (10)$$

Note that

$$(\mathbf{\Lambda} + \mathbf{D}^H \mathbf{D})^{-1} = \mathbf{\Lambda}^{-\frac{1}{2}} (\mathbf{I} + \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{D}^H \mathbf{D} \mathbf{\Lambda}^{-\frac{1}{2}})^{-1} \mathbf{\Lambda}^{-\frac{1}{2}} \quad (11)$$

Using a standard approach, low-rank updates and normalized dictionaries are incompatible. However, with the above modification, the scaling factors (to accomodate the low-rank updates) can be represented in the $\mathbf{\Lambda}$ matrix, leaving the dictionary \mathbf{D} appropriately normalized.

This allows the individual rescaling of filters, but not channels, so individual channels will not be normalized.

The same approach can also be applied in the context of mask decoupling and/or ℓ -1 data fidelity.