Theorem 1. *The iterative algorithm for solving the composite regularization problem*

$$\min_{\boldsymbol{x}} \left\{ \mathcal{J}(\boldsymbol{x}) = \underbrace{\frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2}_{f(\boldsymbol{x})} + \underbrace{\sum_{i=1}^{3} \alpha_i g_i(\boldsymbol{x})}_{g(\boldsymbol{x})} \right\}, \quad \text{where } \sum_{i=1}^{3} \alpha_i = 1,\ 0 < \alpha_i\ \forall i, \tag{1}$$

can be unrolled into a deep-unfolded network with the feedforward operation

$$\boldsymbol{x}^{(l+1)} = \sum_{i=1}^{3} \alpha_i^{(l)} \mathcal{P}_i \left(\mathbf{S}\boldsymbol{x}^{(l)} + \mathbf{W}\boldsymbol{y}; \Theta_i^{(l)} \right),$$

where $\mathbf{S}$, $\mathbf{W}$, $\{\alpha_i^{(l)}\}$, *and* $\{\Theta_i^{(l)}\}$ *are learnable parameters of the network with* l *representing the layer of the network.*

Proof. The optimization strategy for solving the composite regularization problem in (1) relies on Majorization-Minimization (MM) [1] and proximal averaging [2–6]. The data-fidelity term $f : \mathbb{R}^n \to (-\infty, \infty]$ is proper, closed, and L-smooth, i.e., $\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2 \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_2$.

Thus, there exists $\eta < 1/L$ such that $\mathcal{J}(\boldsymbol{x})$ is upper-bounded locally by a strongly convex function about $\boldsymbol{x} = \boldsymbol{x}^{(l)}$ as follows:

$$\mathcal{J}(\boldsymbol{x}) \leq Q^{(l)}(\boldsymbol{x}), \tag{2}$$

where $Q^{(l)}(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{(l)})^\intercal(\frac{1}{\eta}\mathbf{I} - \mathbf{A}^\intercal\mathbf{A})(\boldsymbol{x} - \boldsymbol{x}^{(l)}) + f(\boldsymbol{x}^{(l)}) + g(\boldsymbol{x})$ at iteration l. Minimizing the majorizer $Q^{(l)}(\boldsymbol{x})$ gives the following update for $\boldsymbol{x}$:

$$\boldsymbol{x}^{(l+1)} = \arg\min_{\boldsymbol{x}} \left\{ \frac{1}{2\eta} \left\| \boldsymbol{x} - \boldsymbol{u}^{(l)} \right\|_2^2 + g(\boldsymbol{x}) \right\}, \tag{3}$$

where $\boldsymbol{u}^{(l)} = \boldsymbol{x}^{(l)} - \eta\mathbf{A}^\intercal(\mathbf{A}\boldsymbol{x}^{(l)} - \boldsymbol{y})$. Since there is no closed-form solution to (3), we employ proximal averaging [4, 5] to obtain an approximate update:

$$\begin{aligned}\boldsymbol{x}^{(l+1)} &= \arg\min_{\boldsymbol{x}} \left\{ \sum_{i=1}^{3} \frac{\alpha_i}{2\eta} \left\| \boldsymbol{x} - \boldsymbol{u}^{(l)} \right\|_2^2 + \sum_{i=1}^{3} \alpha_i g_i(\boldsymbol{x}) \right\}, \\ &= \sum_{i=1}^{3} \alpha_i \mathcal{P}_i(\boldsymbol{u}^{(l)}). \end{aligned} \tag{4}$$

The iterative update in (4) involves linear transformations followed by a nonlinear activation, which can be interpreted as a layer in a feedforward neural network, and is thus unfolded to obtain a network with the following input-output relation for the l^{th} layer:

$$\boldsymbol{x}^{(l+1)} = \sum_{i=1}^{3} \alpha_i^{(l)} \mathcal{P}_i \left(\mathbf{S}\boldsymbol{x}^{(l)} + \mathbf{W}\boldsymbol{y}; \Theta_i^{(l)} \right), \tag{5}$$

where $\mathbf{S}$ is initialized with $\mathbf{I} - \eta\mathbf{A}^\intercal\mathbf{A}$ and $\mathbf{W}$ with $\eta\mathbf{A}^\intercal$. □

1. REFERENCES

[1] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization–minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2980–2991, 2007.

[2] W. Zhong and J. Kwok, "Gradient descent with proximal average for nonconvex and composite regularization," *in Proceedings of AAAI Conference on Artificial Intelligence*, vol. 28, no. 1, 2014.

[3] U. S. Kamilov, "A parallel proximal algorithm for anisotropic total variation minimization," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 539–548, 2016.

[4] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, "The proximal average: basic theory," *SIAM Journal on Optimization*, vol. 19, no. 2, pp. 766–785, 2008.

[5] Y. Yu, "Better approximation and faster algorithm using the proximal average," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 26, pp. 458–466, 2013.

[6] S. Mache, P. K. Pokala, K. Rajendran, and C. S. Seelamantula, "*Nu*SPAN: A proximal average network for nonuniform sparse model — application to seismic reflectivity inversion," 2021, arXiv:2105.00003 [Online].