Supplementary Material for (H)GIR-MRF

Yuli Sun, Lin Lei, Xiang Tan, Dongdong Guan, Junzheng Wu, Gangyao Kuang

I. Derivation of Algorithm 1

Rewrite the minimization model of structured graph learning as follows:

$$\begin{aligned}&\min_{\mathbf{W}} Tr\left(\mathbf{W}^T\mathbf{D}^x\right)+\left\|\mathbf{W}\boldsymbol{\alpha}^{\frac{1}{2}}\right\|_F^2+\beta g\left(\boldsymbol{\varepsilon}\right)\\ &\quad s.t.\ \mathbf{W}\geq 0,\ \mathbf{W}^T\mathbf{1}_{N_S}=\mathbf{1}_{N_S},\ \mathbf{X}=\mathbf{X}\mathbf{W}+\boldsymbol{\varepsilon}.\end{aligned}\tag{1}$$

Problem (1) can be efficiently solved by using the alternating direction method of multipliers (ADMM). First, we introduce an auxiliary variable $\mathbf{S}\in\mathbb{R}^{N_S\times N_S}$, and rewrite the model (1) as the minimization of

$$\begin{aligned}&\mathcal{L}\left(\boldsymbol{\varepsilon},\mathbf{S},\mathbf{W},\mathbf{R}_1,\mathbf{R}_2\right)=Tr\left(\mathbf{W}^T\mathbf{D}^x\right)+\left\|\mathbf{W}\boldsymbol{\alpha}^{\frac{1}{2}}\right\|_F^2+\beta g\left(\boldsymbol{\varepsilon}\right)\\ &\quad+Tr\left(\mathbf{R}_1^T\left(\mathbf{W}-\mathbf{S}\right)\right)+Tr\left(\mathbf{R}_2^T\left(\mathbf{X}-\mathbf{X}\mathbf{S}-\boldsymbol{\varepsilon}\right)\right)\\ &\quad+\frac{\mu_1}{2}\left\|\mathbf{W}-\mathbf{S}\right\|_F^2+\frac{\mu_2}{2}\left\|\mathbf{X}-\mathbf{X}\mathbf{S}-\boldsymbol{\varepsilon}\right\|_F^2\\ &s.t.\ \mathbf{W}\geq 0,\ \mathbf{W}^T\mathbf{1}_{N_S}=\mathbf{1}_{N_S},\end{aligned}\tag{2}$$

where $\mathbf{R}_1\in\mathbb{R}^{N_S\times N_S}$ and $\mathbf{R}_2\in\mathbb{R}^{3C_{\mathbf{X}}\times N_S}$ are two Lagrangian multipliers, and $\mu_1,\mu_2>0$ are two penalty parameters. Then the alternating direction method (ADM) can be used to solve the minimization of (2) by iteratively updating one variable at a time and fixing the others. ADM separates (2) into $\boldsymbol{\varepsilon}$-subproblem, $\mathbf{S}$-subproblem and $\mathbf{W}$-subproblem.

$\boldsymbol{\varepsilon}$-subproblem: given the current points $\left(\boldsymbol{\varepsilon}^{(t)},\mathbf{S}^{(t)},\mathbf{W}^{(t)},\mathbf{R}_1^{(t)},\mathbf{R}_2^{(t)}\right)$ at the t-th iteration, the minimization of (2) with respect to $\boldsymbol{\varepsilon}$ can be formulated as follows by fixing the unrelated terms

$$\begin{aligned}\boldsymbol{\varepsilon}^{(t+1)}=\arg\min_{\boldsymbol{\varepsilon}}&\beta g\left(\boldsymbol{\varepsilon}\right)+Tr\left(\left(\mathbf{R}_2^{(t)}\right)^T\left(\mathbf{X}-\mathbf{X}\mathbf{S}^{(t)}-\boldsymbol{\varepsilon}\right)\right)\\ &+\frac{\mu_2}{2}\left\|\mathbf{X}-\mathbf{X}\mathbf{S}^{(t)}-\boldsymbol{\varepsilon}\right\|_F^2.\end{aligned}\tag{3}$$

Define the proximal operator as

$$prox_{\lambda g}\left(\mathbf{Y}\right):=\arg\min_{\mathbf{X}} g\left(\mathbf{X}\right)+\frac{1}{2\lambda}\left\|\mathbf{X}-\mathbf{Y}\right\|_F^2.\tag{4}$$

We have that $\boldsymbol{\varepsilon}^{(t+1)}$ can be obtained by

$$\boldsymbol{\varepsilon}^{(t+1)}=prox_{\frac{\beta}{\mu_2}g}\left(\mathbf{Q}^{(t)}\right),\tag{5}$$

where $\mathbf{Q}^{(t)}=\mathbf{X}-\mathbf{X}\mathbf{S}^{(t)}+\frac{\mathbf{R}_2^{(t)}}{\mu_2}$. According to different regularization strategies, $\boldsymbol{\varepsilon}^{(t+1)}$ can be updated with different closed-form solutions of (5). When we select the squared Frobenius norm $g\left(\boldsymbol{\varepsilon}\right)=\left\|\boldsymbol{\varepsilon}\right\|_F^2$, the $\boldsymbol{\varepsilon}^{(t+1)}$ can be computed by

$$\boldsymbol{\varepsilon}^{(t+1)}=\frac{\mathbf{Q}^{(t)}}{1+2\beta/\mu_2}.\tag{6}$$

When we select the ℓ_1-norm $g\left(\boldsymbol{\varepsilon}\right)=\left\|\boldsymbol{\varepsilon}\right\|_1$, the $\boldsymbol{\varepsilon}^{(t+1)}$ can be updated as

$$\boldsymbol{\varepsilon}^{(t+1)}=shrink\left\{\mathbf{Q}^{(t)},\frac{\beta}{\mu_2}\right\},\tag{7}$$

where $shrink\left\{\mathbf{X},\lambda\right\}$ is an element-wise soft shrinkage operator defined as

$$\left[shrink\left\{\mathbf{X},\lambda\right\}\right]_{i,j}=sign\left(x_{i,j}\right)\max\left\{\left|x_{i,j}\right|-\lambda,0\right\}.\tag{8}$$

When we select the $\ell_{2,1}$-norm $g\left(\boldsymbol{\varepsilon}\right)=\left\|\boldsymbol{\varepsilon}\right\|_{2,1}$, we can update $\boldsymbol{\varepsilon}^{(t+1)}$ by using the Lemma 3.3 of [1]

$$\boldsymbol{\varepsilon}_i^{(t+1)}=\max\left\{\left\|\mathbf{Q}_i^{(t)}\right\|_2-\frac{\beta}{\mu_2},0\right\}\frac{\mathbf{Q}_i^{(t)}}{\left\|\mathbf{Q}_i^{(t)}\right\|_2},\tag{9}$$

where we follow the convention $0\cdot(0/0)=0$.

S-subproblem: with the fixed points $\left(\boldsymbol{\varepsilon}^{(t+1)},\mathbf{W}^{(t)},\mathbf{R}_1^{(t)},\mathbf{R}_2^{(t)}\right)$, we can update $\mathbf{S}^{(t+1)}$ by taking derivative of $L\left(\boldsymbol{\varepsilon},\mathbf{S},\mathbf{W},\mathbf{R}_1,\mathbf{R}_2\right)$ with respect to $\mathbf{S}$ with the unrelated term fixed and setting it to zero, then we obtain

$$\begin{aligned}\mathbf{S}^{(t+1)}&=\arg\min_{\mathbf{S}}-Tr\left(\left(\mathbf{R}_1^{(t)}\right)^T\mathbf{S}\right)-Tr\left(\left(\mathbf{R}_2^{(t)}\right)^T\mathbf{X}\mathbf{S}\right)\\ &\qquad+\frac{\mu_1}{2}\left\|\mathbf{W}^{(t)}-\mathbf{S}\right\|_F^2+\frac{\mu_2}{2}\left\|\mathbf{X}-\mathbf{X}\mathbf{S}-\boldsymbol{\varepsilon}^{(t+1)}\right\|_F^2\\ &=\left(\mathbf{I}_{N_S}+\frac{\mu_2}{\mu_1}\mathbf{X}^T\mathbf{X}\right)^{-1}\boldsymbol{\Phi}^{(t+1)},\end{aligned}\tag{10}$$

where $\mathbf{\Phi}^{(t+1)} = \mathbf{W}^{(t)} + \frac{\mu_2 \mathbf{X}^T\left(\mathbf{X} - \boldsymbol{\varepsilon}^{(t+1)}\right)}{\mu_1} + \frac{\mathbf{R}_1^{(t)} + \mathbf{X}^T \mathbf{R}_2^{(t)}}{\mu_1}$.

W-subproblem: with the fixed points $\left(\boldsymbol{\varepsilon}^{(t+1)}, \mathbf{S}^{(t+1)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}\right)$, the minimization of (2) with respect to $\mathbf{W}$ can be formulated as

$$\begin{aligned} \mathbf{W}_i^{(t+1)} = & \arg\min_{\mathbf{W}_i} \left(\mathbf{D}_i^x + \mathbf{R}_{1i}^{(t)}\right)^T \mathbf{W}_i + \alpha_i \left\|\mathbf{W}_i\right\|_2^2 \\ & + \frac{\mu_1}{2}\left\|\mathbf{W}_i - \mathbf{S}_i^{(t+1)}\right\|_2^2 \\ & s.t.\ \mathbf{W}_i > 0,\ {\mathbf{W}_i}^T \mathbf{1}_{N_S} = 1, \end{aligned} \tag{11}$$

where $\mathbf{R}_{1i}^{(t)}$ is the i-th column of $\mathbf{R}_1^{(t)}$. By defining $\gamma_i = \frac{\mu_1}{2} + \alpha_i$ and $\mathbf{P}_i = \mathbf{D}_i^x + \mathbf{R}_{1i}^{(t)} - \mu_1 \mathbf{S}_i^{(t+1)}$, the Lagrangian function of (11) is

$$\begin{aligned} \mathcal{L}\left(\mathbf{W}_i, \varsigma, \boldsymbol{\upsilon}\right) = & \gamma_i \left\|\mathbf{W}_i\right\|_2^2 + \mathbf{P}_i^T \mathbf{W}_i \\ & - \varsigma\left({\mathbf{W}_i}^T \mathbf{1}_{N_S} - 1\right) - \boldsymbol{\upsilon}^T \mathbf{W}_i, \end{aligned} \tag{12}$$

where $\varsigma \in \mathbb{R}$ and $\boldsymbol{\upsilon} \in \mathbb{R}^{N_S}$ are two Lagrangian multipliers. By using the Karush-Kuhn-Tucker (KKT) condition, the closed-form solution of $\mathbf{W}_i^{(t+1)}$ is

$$\mathbf{W}_i^{(t+1)} = \left(\frac{-\mathbf{P}_i + \varsigma}{2\gamma_i}\right)_+, \tag{13}$$

where the operator $(\mathbf{X})_+$ turns negative elements in $\mathbf{X}$ to zero while keep the rest. Then, we can find that $\mathbf{W}_i^{(t+1)}$ is a sparse vector. We assume that $\mathbf{W}_i^{(t+1)}$ have k_i nonzero elements, i.e., only k_i neighbors have a connection with $\tilde{\mathbf{X}}_i$ with probabilities $\mathbf{W}_i^{(t+1)}$. We sort $\mathbf{P}_i$ in ascending order denoted as $P_{(1),i}, P_{(2),i}, \cdots, P_{(N_S),i}$. Then, the following inequalities hold

$$\left\{ \begin{array}{l} \frac{-P_{(k_i),i} + \varsigma}{2\gamma_i} > 0 \\ \frac{-P_{(k_i+1),i} + \varsigma}{2\gamma_i} \leq 0 \end{array} \right. . \tag{14}$$

Considering the constraint ${\mathbf{W}_i}^T \mathbf{1}_{N_S} = 1$, we have

$$\sum_{h=1}^{k_i} \frac{-P_{(h),i} + \varsigma}{2\gamma_i} = 1. \tag{15}$$

Substitute (15) into (14), we have

$$\left\{ \begin{array}{l} \gamma_i > \frac{k_i}{2} P_{(k_i),i} - \frac{1}{2}\sum_{h=1}^{k_i} P_{(h),i} \\ \gamma_i \leq \frac{k_i}{2} P_{(k_i+1),i} - \frac{1}{2}\sum_{h=1}^{k_i} P_{(h),i} \end{array} \right. . \tag{16}$$

Then, we can find that the regularization parameter α_i can be replaced by number of neighbors k_i when we set

$$\alpha_i = \frac{k_i}{2} P_{(k_i+1),i} - \frac{1}{2}\sum\nolimits_{h=1}^{k_i} P_{(h),i} - \frac{\mu_1}{2}. \tag{17}$$

Therefore, the tuning of parameter α_i becomes the tuning of k_i, which is more intuitive (it has explicit meaning) and easier (it is an integer). With the k_i, the $\mathbf{W}_i^{(t+1)}$ can be updated as

$$W_{(j),i}^{(t+1)} = \left\{ \begin{array}{ll} \frac{P_{(k_i+1),i} - P_{(j),i}}{k_i P_{(k_i+1),i} - \sum_{h=1}^{k_i} P_{(h),i}}, & j \leq k_i \\ 0, & j > k_i \end{array} \right. . \tag{18}$$

Multipliers: finally, the two Lagrangian multipliers can be updated as

$$\begin{aligned} \mathbf{R}_1^{(t+1)} &= \mathbf{R}_1^{(t)} + \mu_1\left(\mathbf{W}^{(t+1)} - \mathbf{S}^{(t+1)}\right), \\ \mathbf{R}_2^{(t+1)} &= \mathbf{R}_2^{(t)} + \mu_2\left(\mathbf{X} - \mathbf{X}\mathbf{S}^{(t+1)} - \boldsymbol{\varepsilon}^{(t+1)}\right). \end{aligned} \tag{19}$$

II. Loop of the graph G

Lemma 1. *The constructed graph $G = (V, E, w)$ has a loop for each vertex, that is, $w_{i,i} > 0$ for $i = 1, \cdots, N_S$.*

Proof: Assume by contradiction that there exists such optimal $\mathbf{W}^*$ which has at least one diagonal element of 0. Without loss of generality, we assume that $w_{i,i}^* = 0$. By the definition, $\mathbf{W}^*$ is the optimal solution of the following minimization problem

$$\begin{aligned} & \min_{\mathbf{W}} \sum_{i=1}^{N_S}\sum_{j=1}^{N_S} \left\|\mathbf{X}_i - \mathbf{X}_j\right\|_2^2 w_{i,j} + \sum_{i=1}^{N_S} \alpha_i \left\|\mathbf{W}_i\right\|_2^2 + \beta g\left(\boldsymbol{\varepsilon}\right) \\ & s.t.\ 0 \leq w_{i,j} \leq 1,\ \sum_{i=1}^{N} w_{i,j} = 1,\ \mathbf{X} = \mathbf{X}\mathbf{W} + \boldsymbol{\varepsilon}, \end{aligned} \tag{20}$$

where $\alpha_i, \beta > 0$ are the balancing parameters, and $g\left(\boldsymbol{\varepsilon}\right)$ represents the penalty term, which can be the $g\left(\boldsymbol{\varepsilon}\right) = \left\|\boldsymbol{\varepsilon}\right\|_F^2$, $\left\|\boldsymbol{\varepsilon}\right\|_1$, or $\left\|\boldsymbol{\varepsilon}\right\|_{2,1}$. Due to the columnwise independence property of $\mathbf{W}$ in (20), we have that $\mathbf{W}_i^*$ is the optimal solution of the following problem

$$\begin{aligned} & \min_{\mathbf{W}_i} \mathbf{D}_i^T \mathbf{W}_i + \alpha_i \left\|\mathbf{W}_i\right\|_2^2 + \beta g\left(\boldsymbol{\varepsilon}_i\right) \\ & s.t.\ 0 \leq w_{j,i} \leq 1,\ \sum_{i=1}^{N} w_{j,i} = 1,\ \mathbf{X}_i = \mathbf{X}\mathbf{W}_i + \boldsymbol{\varepsilon}_i, \end{aligned} \tag{21}$$

where $\mathbf{D}_i$ is a distance vector with j-th element being $d_{j,i} = \left\|\mathbf{X}_i - \mathbf{X}_j\right\|_2^2$. Denote the objective function of (21) as

$$\mathbf{\Phi}\left(\mathbf{W}_i\right) = \mathbf{D}_i^T \mathbf{W}_i + \alpha_i \left\|\mathbf{W}_i\right\|_2^2 + \beta g\left(\boldsymbol{\varepsilon}_i\right). \tag{22}$$

First, we construct a new $\bar{\mathbf{W}}_i$ as

$$\bar{w}_{j,i} = \left\{ \begin{array}{ll} \delta w_{j,i}^*, & \text{if } j \neq i \\ 1 - \delta, & \text{if } j = i \end{array} \right. , \tag{23}$$

with

$$\delta = \frac{1 - \|\mathbf{W}_i^*\|_2^2}{1 + \|\mathbf{W}_i^*\|_2^2} + \zeta, \tag{24}$$

and ζ is an arbitrary positive constant that satisfies $\zeta \in \left(0, \frac{2\|\mathbf{W}_i^*\|_2^2}{1+\|\mathbf{W}_i^*\|_2^2}\right)$, which can enable $\frac{1-\|\mathbf{W}_i^*\|_2^2}{1+\|\mathbf{W}_i^*\|_2^2} < \delta < 1$. Then we have $0 \leq \bar{w}_{j,i} \leq 1$ and

$$\sum_{j=1}^{N_S} \bar{w}_{j,i} = \bar{w}_{i,i} + \sum_{j \neq i} \delta w_{j,i}^* = 1 - \delta + \sum_{j=1}^{N_S} \delta w_{j,i}^* = 1, \tag{25a}$$

$$\begin{aligned}\mathbf{D}_i^T \bar{\mathbf{W}}_i &= \sum_{j=1}^{N_S} \|\mathbf{X}_i - \mathbf{X}_j\|_2^2 \bar{w}_{j,i} = \sum_{j \neq i} \delta \|\mathbf{X}_i - \mathbf{X}_j\|_2^2 w_{j,i}^* \\ &= \delta \mathbf{D}_i^T \mathbf{W}_i^*,\end{aligned} \tag{25b}$$

$$\left\|\bar{\mathbf{W}}_i\right\|_2^2 = \bar{w}_{i,i}^2 + \sum_{j \neq i} \left(\delta w_{j,i}^*\right)^2 = (1-\delta)^2 + \delta^2 \|\mathbf{W}_i^*\|_2^2, \tag{25c}$$

$$\begin{aligned}\bar{\boldsymbol{\varepsilon}}_i &= \mathbf{X}_i - \mathbf{X}\bar{\mathbf{W}}_i = \mathbf{X}_i - \mathbf{X}_i \bar{w}_{i,i} - \sum_{j \neq i} \delta \mathbf{X}_j w_{j,i}^* \\ &= \delta \left(\mathbf{X}_i - \sum_{j=1}^{N_S} \mathbf{X}_j w_{j,i}^* \right) = \delta \boldsymbol{\varepsilon}_i^*.\end{aligned} \tag{25d}$$

Substitute (25) into (22), we have

$$\begin{aligned}\boldsymbol{\Phi}\left(\bar{\mathbf{W}}_i\right) =& \delta \mathbf{D}_i^T \mathbf{W}_i^* + \alpha_i \left((1-\delta)^2 + \delta^2 \|\mathbf{W}_i^*\|_2^2 \right) \\ &+ \beta g\left(\delta \boldsymbol{\varepsilon}_i^*\right).\end{aligned} \tag{26}$$

Then, we have

$$\begin{aligned}&\boldsymbol{\Phi}\left(\mathbf{W}_i^*\right) - \boldsymbol{\Phi}\left(\bar{\mathbf{W}}_i\right) \\ &= (1-\delta)\mathbf{D}_i^T \mathbf{W}_i^* + \beta\left(g\left(\boldsymbol{\varepsilon}_i^*\right) - g\left(\delta\boldsymbol{\varepsilon}_i^*\right)\right) \\ &\quad + \alpha_i \left(\left(1-\delta^2\right) \left\|\bar{\mathbf{W}}_i\right\|_2^2 - (1-\delta)^2 \right) \\ &= (1-\delta)\mathbf{D}_i^T \mathbf{W}_i^* + \beta\left(g\left(\boldsymbol{\varepsilon}_i^*\right) - g\left(\delta\boldsymbol{\varepsilon}_i^*\right)\right) \\ &\quad + \alpha_i (1-\delta) \left(\delta \left(1 + \|\mathbf{W}_i^*\|_2^2\right) - \left(1 - \|\mathbf{W}_i^*\|_2^2\right) \right) \\ &\geq \alpha_i (1-\delta) \left(\delta \left(1 + \|\mathbf{W}_i^*\|_2^2\right) - \left(1 - \|\mathbf{W}_i^*\|_2^2\right) \right) \\ &> 0,\end{aligned} \tag{27}$$

where the first inequality comes from $\mathbf{D}_i^T \mathbf{W}_i^* \geq 0$ and $g\left(\boldsymbol{\varepsilon}_i^*\right) \geq g\left(\delta\boldsymbol{\varepsilon}_i^*\right)$, and the second inequality comes from (24). Then, we have $\boldsymbol{\Phi}\left(\mathbf{W}_i^*\right) > \boldsymbol{\Phi}\left(\bar{\mathbf{W}}_i\right)$, which contradicts the optimality of $\mathbf{W}_i^*$. Then, we have $w_{i,i}^* > 0$. This completes the proof. ∎

III. Reformulation of the HGLR

Reformulate the HGLR as follows:

$$\begin{aligned}&\sum_{e \in E^h} \sum_{\left\{\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_j\right\} \in e} \frac{w^h(e)\, \tilde{h}\left(\tilde{\mathbf{X}}_i, e\right) \tilde{h}\left(\tilde{\mathbf{X}}_j, e\right)}{\delta(e)} \|\mathbf{Z}_i - \mathbf{Z}_j\|_2^2 \\ &= \sum_{\tilde{\mathbf{X}}_i \in V^h} \mathbf{Z}_i^T \mathbf{Z}_i \sum_{e \in E^h} w^h(e)\, \tilde{h}\left(\tilde{\mathbf{X}}_i, e\right) \sum_{\tilde{\mathbf{X}}_j \in V^h} \frac{\tilde{h}\left(\tilde{\mathbf{X}}_j, e\right)}{\delta(e)} \\ &\quad + \sum_{\tilde{\mathbf{X}}_j \in V^h} \mathbf{Z}_j^T \mathbf{Z}_j \sum_{e \in E^h} w^h(e)\, \tilde{h}\left(\tilde{\mathbf{X}}_j, e\right) \sum_{\tilde{\mathbf{X}}_i \in V^h} \frac{\tilde{h}\left(\tilde{\mathbf{X}}_i, e\right)}{\delta(e)} \\ &\quad - \sum_{e \in E^h} \sum_{\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_j \in V^h} 2\mathbf{Z}_i^T \frac{w^h(e)\, \tilde{h}\left(\tilde{\mathbf{X}}_i, e\right) \tilde{h}\left(\tilde{\mathbf{X}}_j, e\right)}{\delta(e)} \mathbf{Z}_j.\end{aligned} \tag{28}$$

Substitute the definitions of $d(v) = \sum_{e \in E^h} w^h(e)\, h(v,e)$, $\delta(e) = \sum_{v \in V^h} h(v,e)$ into (28), we have

$$\begin{aligned}&\sum_{e \in E^h} \sum_{\left\{\tilde{\mathbf{X}}_i, \tilde{\mathbf{X}}_j\right\} \in e} \frac{w^h(e)\, \tilde{h}\left(\tilde{\mathbf{X}}_i, e\right) \tilde{h}\left(\tilde{\mathbf{X}}_j, e\right)}{\delta(e)} \|\mathbf{Z}_i - \mathbf{Z}_j\|_2^2 \\ &= 2Tr\left(\mathbf{Z}\mathbf{D}_v\mathbf{Z}^T\right) - 2Tr\left(\mathbf{Z}\tilde{\mathbf{H}}\mathbf{W}^h\mathbf{D}_e^{-1}\tilde{\mathbf{H}}^T\mathbf{Z}^T\right) \\ &= 2Tr\left(\mathbf{Z}\mathbf{L}^h\mathbf{Z}^T\right),\end{aligned} \tag{29}$$

where the last equation comes from the definition of hypergraph Laplacian matrix as $\mathbf{L}^h = \mathbf{D}_v - \tilde{\mathbf{H}}\mathbf{W}^h\mathbf{D}_e^{-1}\tilde{\mathbf{H}}^T$.

References

[1] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.