Supplementary Material for "Multivariate regression modeling in integrative analysis via sparse regularization"

by Shuichi Kawano, Toshikazu Fukushima, Junichi Nakagawa, Mamoru Oshiki

S1 Computational algorithm of the extended model

The minimization problem

$$\min_{\substack{\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^M\\ B^1,\ldots,B^M\\ C^1,\ldots,C^M\\ \Sigma^1,\ldots,\Sigma^M}} \left[\sum_{m=1}^{M} \left\{ \frac{1}{2n_m} \left\| Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - X^m B^m - Z^m C^m \right\|^2_{(\Sigma^m)^{-1}} - \frac{1}{2}\log\det\left((\Sigma^m)^{-1}\right) + \frac{\nu_m}{2}\mathrm{tr}\left((\Sigma^m)^{-1}\right) \right\} \right.$$
$$\left. + \sum_{j=1}^{p}\sum_{k=1}^{q} \lambda_{jk}\|\boldsymbol{\beta}_{jk}\|_2 + \sum_{m=1}^{M}\gamma_m\|C^m\|_1 \right],$$

is rewritten as

$$\min_{\substack{\boldsymbol{\alpha}^1,\ldots,\boldsymbol{\alpha}^M\\ B^1,\ldots,B^M\\ C^1,\ldots,C^M\\ \Sigma^1,\ldots,\Sigma^M\\ \boldsymbol{\eta}_{11},\ldots,\boldsymbol{\eta}_{pq}\\ D^1,\ldots,D^M}} \left[\sum_{m=1}^{M} \left\{ \frac{1}{2n_m} \left\| Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - X^m B^m - Z^m C^m \right\|^2_{(\Sigma^m)^{-1}} - \frac{1}{2}\log\det\left((\Sigma^m)^{-1}\right) + \frac{\nu_m}{2}\mathrm{tr}\left((\Sigma^m)^{-1}\right) \right\} \right.$$
$$\left. + \sum_{j=1}^{p}\sum_{k=1}^{q} \lambda_{jk}\|\boldsymbol{\eta}_{jk}\|_2 + \sum_{m=1}^{M}\gamma_m\|D^m\|_1 \right]$$
$$\text{subject to} \quad \boldsymbol{\eta}_{jk} = \boldsymbol{\beta}_{jk}, \quad C^m = D^m \tag{S.1}$$

From the problem (S.1), we can obtain the augmented Lagrangian in the form

$$\sum_{m=1}^{M} \left\{ \frac{1}{2n_m} \left\| Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - X^m B^m - Z^m C^m \right\|^2_{(\Sigma^m)^{-1}} - \frac{1}{2}\log\det\left((\Sigma^m)^{-1}\right) + \frac{\nu_m}{2}\mathrm{tr}\left((\Sigma^m)^{-1}\right) \right\}$$
$$+ \sum_{j=1}^{p}\sum_{k=1}^{q} \lambda_{jk}\|\boldsymbol{\eta}_{jk}\|_2 + \sum_{m=1}^{M}\gamma_m\|D^m\|_1 + \frac{\rho}{2}\sum_{j=1}^{p}\sum_{k=1}^{q}\|\boldsymbol{\eta}_{jk} - \boldsymbol{\beta}_{jk} + \boldsymbol{u}_{jk}\|_2^2 + \frac{\rho}{2}\sum_{m=1}^{M}\|C^m - D^m + V^m\|_F^2$$

where $\boldsymbol{u}_{jk}$ $(j = 1,\ldots,p;\ k = 1,\ldots,q)$ and V^m $(m = 1,\ldots,M)$ are dual variables and ρ is a penalty parameter with a positive value.

Thus, the update equations by ADMM are, respectively, given as follows.

Update of $\boldsymbol{\eta}_{jk}$:

$$\boldsymbol{\eta}_{jk} = \mathcal{S}\left(\boldsymbol{\beta}_{jk} - \boldsymbol{u}_{jk}, \frac{\lambda_{jk}}{\rho}\right),$$

where $\mathcal{S}(\cdot,\cdot)$ is the soft-thresholding operator given by

$$\mathcal{S}(\boldsymbol{a}, b) = \left(1 - \frac{b}{\|\boldsymbol{a}\|_2}\right)_+ \boldsymbol{a}.$$

Update of B^m: We set H^m having η_{jk}^m as the (j,k)-th element and U^m having u_{jk}^m as the (j,k)-th element. Since $\sum_{j=1}^p \sum_{k=1}^q \|\boldsymbol{\eta}_{jk} - \boldsymbol{\beta}_{jk} + \boldsymbol{u}_{jk}\|_2^2 = \sum_{m=1}^M \|B^m - H^m - U^m\|_F^2$, we consider the quantity

$$\sum_{m=1}^M \frac{1}{2n_m} \left\|Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - X^m B^m - Z^m C^m\right\|_{(\Sigma^m)^{-1}}^2 + \frac{\rho}{2} \sum_{m=1}^M \|B^m - H^m - U^m\|_F^2.$$

From this quantity, we obtain

$$\begin{aligned}\mathrm{vec}(B^m) = &\left\{(\Sigma^m)^{-1} \otimes (X^m)^\top X^m + n_m \rho I_{pq}\right\}^{-1} \\ &\mathrm{vec}\left[(X^m)^\top \left\{Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - Z^m C^m\right\} (\Sigma^m)^{-1} + n_m \rho (H^m + U^m)\right],\end{aligned}$$

where I_p is a $p \times p$ identity matrix.

Update of D^m: We consider

$$\rho \sum_{m=1}^M \left[\frac{1}{2}\|D^m - (C^m + V^m)\|_F^2 + \frac{\gamma_m}{\rho}\|D^m\|_1\right].$$

From this quantity, we obtain

$$(D^m)_{ij} = S\left((C^m + V^m)_{ij}, \frac{\gamma_m}{\rho}\right), \quad (i = 1, \ldots, r_m;\ j = 1, \ldots, q),$$

where $S(\cdot,\cdot)$ is the soft-thresholding operator defined by $S(a,b) = \mathrm{sign}(a)(|a| - b)_+$ and $(A)_{ij}$ denotes an (i,j)-th element of a matrix A.

Update of C^m: We consider the quantity

$$\sum_{m=1}^M \frac{1}{2n_m} \left\|Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - X^m B^m - Z^m C^m\right\|_{(\Sigma^m)^{-1}}^2 + \frac{\rho}{2} \sum_{m=1}^M \|C^m - D^m + V^m\|_F^2.$$

From this quantity, we obtain

$$\begin{aligned}\mathrm{vec}(C^m) = &\left\{(\Sigma^m)^{-1} \otimes (Z^m)^\top Z^m + n_m \rho I_{qr_m}\right\}^{-1} \\ &\mathrm{vec}\left[(Z^m)^\top \left\{Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - X^m B^m\right\} (\Sigma^m)^{-1} + n_m \rho (D^m - V^m)\right].\end{aligned}$$

Update of $\boldsymbol{\alpha}^m$:

$$\boldsymbol{\alpha}^m = \frac{1}{n_m}(Y^m - X^m B^m - Z^m C^m)^\top \mathbf{1}_{n_m}.$$

Update of $\boldsymbol{\Sigma}^m$: From MLE of multivariate regression models, we obtain

$$\Sigma^m = \frac{1}{n_m}\left\{Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - X^m B^m - Z^m C^m\right\}^\top \left\{Y^m - \mathbf{1}_{n_m}(\boldsymbol{\alpha}^m)^\top - X^m B^m - Z^m C^m\right\} + \nu_m I_q,$$

where I_q is a $q \times q$ identity matrix.

Update of $\boldsymbol{u}_{jk}$:

$$\boldsymbol{u}_{jk} = \boldsymbol{u}_{jk} + \boldsymbol{\eta}_{jk} - \boldsymbol{\beta}_{jk}, \quad (j = 1, \ldots, p;\ k = 1, \ldots, q).$$

Update of V^m:

$$V^m = V^m + C^m - D^m, \quad (m = 1, \ldots, M).$$

S2 List of selected species

Table S.1: Selected species of MRR related to $NO2^-$ in Dataset 1.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo7544	Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae		

Table S.2: Selected species of MRR related to SCN^- in Dataset 1.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo791	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo3189	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo4985	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo6554	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Halothiobacillaceae	Halothiobacillus	Halothiobacillus hydrothermalis
denovo7058	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae		
denovo8965	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae		

Table S.3: Selected species of MRR related to $NO2^-$ in Dataset 2.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo7544	Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae		

Table S.4: Selected species of MRR related to SCN^- in Dataset 2.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo167	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo630	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo791	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo3546	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo4985	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo6554	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Halothiobacillaceae	Halothiobacillus	Halothiobacillus hydrothermalis
denovo7350	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo8965	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae		
denovo10281	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo10814	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo15497	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Halothiobacillaceae	Halothiobacillus	

Table S.5: Selected species of UR related to $NO2^-$ in Dataset 1.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo8384	Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae		

Table S.6: Selected species of UR related to SCN^- in Dataset 1.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo3189	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo7058	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae		
denovo12987	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			

Table S.7: Selected species of UR related to $NO2^-$ in Dataset 2.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo8384	Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae		

Table S.8: Selected species of UR related to SCN^- in Dataset 2.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo167	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo630	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo3546	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo7350	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo10281	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo10814	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales			
denovo15497	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Halothiobacillaceae	Halothiobacillus	

Table S.9: Selected species of lasso related to $NO2^-$ in Dataset 1.

OTU ID	Domain	Phylum	Class	Order	Family	Genus	Species
denovo8384	Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae		