AN ASSEMBLY AND DECOMPOSITION APPROACH FOR CONSTRUCTING SEPARABLE MINORIZING FUNCTIONS IN A CLASS OF MM ALGORITHMS

Guo-Liang TIAN[†], Xi-Fen HUANG[‡*] and Jinfeng XU[‡]

[†]*Southern University of Science and Technology*

[‡*]*Yunnan University of Finance and Economics*

[‡]*The University of Hong Kong*

S1 The proofs of Proposition 1 and Proposition 2

Proof of Proposition 1. Since $\nabla Q[M(\boldsymbol{\theta})|\boldsymbol{\theta}] = \mathbf{0}$, it is easy to show that $M(\boldsymbol{\theta})$ is continuously differentiable with differential

$$dM(\boldsymbol{\theta}) = -d^{20}Q[M(\boldsymbol{\theta})|\boldsymbol{\theta}]^{-1}d^{11}Q[M(\boldsymbol{\theta})|\boldsymbol{\theta}]. \tag{S1.1}$$

Furthermore, $\nabla\ell(\boldsymbol{\theta}) - \nabla Q[M(\boldsymbol{\theta})|\boldsymbol{\theta}] = \mathbf{0}$. Taking differential on both sides and set $\boldsymbol{\theta} = \boldsymbol{\theta}^{\infty}$, we have

$$d^2\ell(\boldsymbol{\theta}^{\infty}) - d^{20}Q(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty}) - d^{11}Q(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty}) = \mathbf{0}. \tag{S1.2}$$

Substituting (S1.2) into (S1.1), we have $dM(\boldsymbol{\theta}^{\infty}) = I - d^{20}Q(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty})^{-1}d^2\ell(\boldsymbol{\theta}^{\infty})$. By Lange's Lemma, it is then sufficient to show that all the eigenvalues of the differential $dM(\boldsymbol{\theta}^{\infty})$ belong to $[0, 1)$. Here we determine the eigenvalues of $dM(\boldsymbol{\theta}^{\infty})$ by the stationary values of the Rayleigh quotient

$$R(v) = \frac{v^{\top}[d^{20}Q(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty}) - d^2\ell(\boldsymbol{\theta}^{\infty})]v}{v^{\top}d^{20}Q(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty})v} = 1 - \frac{v^{\top}d^2\ell(\boldsymbol{\theta}^{\infty})v}{v^{\top}d^{20}Q(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty})v}.$$

At the optimal point $\boldsymbol{\theta}^{\infty}$, both $d^2\ell(\boldsymbol{\theta}^{\infty})$ and $Q(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty})$ are negative definite and $R(v) < 1$ for any unit vector v. The maximum of $R(v)$ is strictly less than 1. Note also that $d^{20}Q(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty}) - d^2\ell(\boldsymbol{\theta}^{\infty})$ is negative semidefinite. It follows that $R(v) \geqslant 0$ and the minimum of $R(v)$ is not less than 0.

Proof of Proposition 2. Let Γ be the set of cluster points generated by the sequence $\boldsymbol{\theta}^{(t+1)} = M(\boldsymbol{\theta}^{(t)})$ starting from the initial value $\boldsymbol{\theta}^{(0)}$. By the

Liapunov's theorem in Lange (2010), Γ is contained in the set Δ of stationary points of $\ell(\boldsymbol{\theta})$. On the other hand, Γ is a closed subset of the compact set $\{\boldsymbol{\theta} \in \Omega : \ell(\boldsymbol{\theta}) \geqslant \ell(\boldsymbol{\theta}^{(0)})\}$ and this implies Γ is also compact. According to Proposition 8.2.1 in Lange (2010), Γ is connected. The condition that all stationary points of $\ell(\boldsymbol{\theta})$ are isolated easily implies that the number of stationary points in the compact set $\{\boldsymbol{\theta} \in \Omega : \ell(\boldsymbol{\theta}) \geqslant \ell(\boldsymbol{\theta}^{(0)})\}$ can only be finite. Since the cluster set Γ is a connected subset of finite set Δ, Γ reduces to a singleton.

S2 The EM algorithm for the CZIGP example

If the random variable $\mathbf{y} \sim \text{CZIGP}_m(\phi_0,\boldsymbol{\phi},\boldsymbol{\lambda},\boldsymbol{\pi})$, we have the following stochastic representation (SR):

$$\mathbf{y} \stackrel{\triangle}{=} Z_0(Z_1X_1^*, \cdots, Z_mX_m^*)^\top,$$

where $\{Z_k\}_{k=0}^m \stackrel{\text{ind}}{\sim} \text{Bernoulli}(1-\phi_k)$, $\{X_i^*\}_{i=1}^m \stackrel{\text{ind}}{\sim} \text{GP}(\lambda_i, \pi_i)$ and $\{Z_k\}_{k=0}^m$ and $\{X_i^*\}_{i=1}^m$ are mutually independent. For each $\mathbf{y}_j = (y_{1j}, \cdots, y_{mj})^T$ with $j \in \{1, \cdots, n\}$, based on the above SR, we introduce independent latent variables

$$Z_{0j} \stackrel{\text{iid}}{\sim} \text{Bernoulli}(1-\phi_0),\ Z_{ij} \stackrel{\text{ind}}{\sim} \text{Bernoulli}(1-\phi_i),\ X_{ij}^* \stackrel{\text{ind}}{\sim} \text{GP}(\lambda_i, \pi_i),$$

for $i = 1, \cdots, m$. We denote the missing data by $Y_{mis} = \{Z_{0j}, \{Z_{ij}, X^*_{ij}\}_{i=1}^m\}_{j=1}^n$ and the complete data by $Y_{com} = \{Y_{obs}, Y_{mis}\}$, where z_{0j}, z_{ij}, x^*_{ij} are the realizations of Z_{0j}, Z_{ij} and X^*_{ij}, respectively. Thus, the complete-data likelihood function is given by

$$\begin{aligned} & L(\theta \big| Y_{com}) \\ =& \prod_{j=1}^n \left\{ \phi_0^{1-z_{0j}} (1-\phi_0)^{z_{0j}} \prod_{i=1}^m \left[\phi_i^{1-z_{ij}} (1-\phi_i)^{z_{ij}} \frac{\lambda_i(\lambda_i + \pi_i x^*_{ij})^{x^*_{ij}-1} e^{-(\lambda_i + \pi_i x^*_{ij})}}{x^*_{ij}!} \right] \right\}, \end{aligned}$$

and the complete-data log-likelihood function $\ell(\theta \big| Y_{com})$ is proportional to

$$\begin{aligned} & \sum_{j=1}^n \left[(1-z_{0j}) \log \phi_0 + z_{0j} \log(1-\phi_0) \right] + \sum_{j=1}^n \sum_{i=1}^m \left[(1-z_{ij}) \log \phi_i \right. \\ & \qquad \left. + z_{ij} \log(1-\phi_i) + \log \lambda_i + (x^*_{ij} - 1) \log(\lambda_i + \pi_i x^*_{ij}) - \lambda_i - \pi_i x^*_{ij} \right]. \end{aligned}$$

The M-step is to calculate the complete-data MLEs, which are given by

$$\left\{ \begin{array}{lcl} \phi_0 & = & \dfrac{n - \sum_{j=1}^n z_{0j}}{n}, \\ \phi_i & = & \dfrac{n - \sum_{j=1}^n z_{ij}}{n}, \\ \lambda_i & = & \dfrac{(1-\pi_i) \sum_{j=1}^n x^*_{ij}}{n}, i = 1, \cdots, m, \end{array} \right. \tag{S2.3}$$

while the complete-data MLE of π_i is the root of the equation:

$$H_i(\pi_i | \lambda_i) = \sum_{j=1}^n \frac{x^{*2}_{ij} - x^*_{ij}}{\lambda_i + \pi_i x^*_{ij}} - \sum_{j=1}^n x^*_{ij} = 0, \quad i = 1, \ldots, m. \tag{S2.4}$$

The E-step is to replace $\{z_{0j}\}_{j=1}^n$, $\{z_{ij}\}_{j=1}^n$, $\{x^*_{ij}\}_{j=1}^n$ and $\left\{ \dfrac{x^{*2}_{ij} - x^*_{ij}}{\lambda_i + \pi_i x^*_{ij}} \right\}_{j=1}^n$ by their conditional expectations which are given by

$$\begin{cases} E(z_{0j}\big|Y_{obs},\theta) = (1-\frac{\phi_0}{\gamma_1})I(\boldsymbol{y}_j=\mathbf{0}) + I(\boldsymbol{y}_j\neq\mathbf{0}), \\ E(z_{ij}\big|Y_{obs},\theta) = \big[\psi_i + \frac{(1-\phi_i-\psi_i)\phi_0}{\gamma_1}\big]I(\boldsymbol{y}_j=\mathbf{0}) + \psi_i I(\boldsymbol{y}_j\neq\mathbf{0})I(y_{ji}=0) \\ \qquad\qquad +I(\boldsymbol{y}_j\neq\mathbf{0})I(y_{ji}\neq 0), \\ E(x^*_{ij}\big|Y_{obs},\theta) = \frac{\lambda_i}{1-\pi_i}\big[1-\psi_i+\frac{\psi_i\phi_0}{\gamma_1}\big]I(\boldsymbol{y}_j=\mathbf{0}) + \frac{(1-\psi_i)\lambda_i}{1-\pi_i}I(\boldsymbol{y}_j\neq\mathbf{0})I(y_{ji}=0) \\ \qquad\qquad +y_{ji}I(\boldsymbol{y}_j\neq\mathbf{0})I(y_{ji}\neq 0), \\ E(\frac{{x^*_{ij}}^2-x^*_{ij}}{\lambda_i+\pi_i x^*_{ij}}\big|Y_{obs},\theta) = \frac{\lambda_i}{1-\pi_i}\left(1-\psi_i+\frac{\psi_i\phi_0}{\gamma_1}\right)I(\boldsymbol{y}_j=\mathbf{0}) \\ \qquad\qquad +\frac{(1-\psi_i)\lambda_i}{1-\pi_i}I(\boldsymbol{y}_j\neq\mathbf{0})I(y_{ji}=0) \\ \qquad\qquad +\frac{y_{ji}^2-y_{ji}}{\lambda_i+\pi_i y_{ji}}I(\boldsymbol{y}_j\neq\mathbf{0})I(y_{ji}\neq 0), \end{cases} \tag{S2.5}$$

where $\psi_i = \frac{(1-\phi_i)e^{-\lambda i}}{\phi_i+(1-\phi_i)e^{-\lambda i}}$.

S3 The derivation of the rate matrix

Poisson model for transmission tomography: The rate matrix of PET via MM algorithm is given by

$$E[dM_{\mathrm{MM}}(\boldsymbol{\theta}^\infty)] = \mathbf{I} - E\left[\frac{d^2Q_{\mathrm{MM}}(\boldsymbol{\theta}^\infty|\boldsymbol{\theta}^\infty)}{n}\right]^{-1} E\left[\frac{d^2\ell(\boldsymbol{\theta}^\infty)}{n}\right], \tag{S3.6}$$

where

$$E\left[\frac{1}{n}\mathrm{d}^2\ell(\theta^\infty)\right] = \left\{\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\pi_j\partial\pi_l}\right]\right\}_{jl},$$

$$\left\{\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\pi_j\partial\pi_l}\right]\right\} = \frac{1}{n}\sum_{i=1}^n\left\{-s_i a_{ij}a_{jl}\exp(-a_i^\top\pi) + \frac{s_i r_i \mathrm{e}^{-a_i^\top\pi}a_{ij}a_{il}}{r_i+s_i\mathrm{e}^{-a_i^\top\pi}}\right\},$$

$j = 1, \ldots, q; l = 1, \ldots, q.$

$$E\left[\frac{1}{n}\mathrm{d}^2 Q_{\mathrm{MM}}(\theta^{\infty}|\theta^{\infty})\right] = \begin{pmatrix} \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial \pi_1^2}\right] & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial \pi_q^2}\right], \end{pmatrix}$$

$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial \pi_j^2}\right] = \frac{1}{n}\sum_{i=1}^{n}\left[-a_{ij}^2 s_i w_{ij}^{-1}\exp(-a_i^{\top}\pi)\right],$$

$j = 1, \cdots, q.$

Left-truncated normal distribution: The rate matrices of LTN via MM and EM algorithms are given by

$$\begin{aligned} E[dM_{\mathrm{MM}}(\boldsymbol{\theta}^{\infty})] &= \mathbf{I} - E\left[\frac{d^2 Q_{\mathrm{MM}}(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty})}{n}\right]^{-1} E\left[\frac{d^2\ell(\boldsymbol{\theta}^{\infty})}{n}\right], \\ E[dM_{\mathrm{EM}}(\boldsymbol{\theta}^{\infty})] &= \mathbf{I} - E\left[\frac{d^2 Q_{\mathrm{EM}}(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty})}{n}\right]^{-1} E\left[\frac{d^2\ell(\boldsymbol{\theta}^{\infty})}{n}\right], \end{aligned} \tag{S3.7}$$

set $\delta = \sigma^2$, we have

$$E\left[\frac{1}{n}\mathrm{d}^2\ell(\theta^{\infty})\right] = \begin{pmatrix} \frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\mu^2}\right] & \frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\mu\partial\delta}\right] \\ \frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\mu\partial\delta}\right] & \frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\delta^2}\right] \end{pmatrix},$$

$$
\begin{aligned}
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\mu^2}\right] &= -\delta^{-1}+\frac{\delta^{-1}\phi^2[(a-\mu)\delta^{-\frac{1}{2}}]}{\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}^2} \\
&+ \frac{\delta^{-1}\phi'[(a-\mu)\delta^{-\frac{1}{2}}]+\frac{1}{2}\delta^{-\frac{3}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]}, \\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\mu\partial\delta}\right] &= \frac{\delta^{-\frac{3}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]}+\frac{(a-\mu)\delta^{-2}\phi^2[(a-\mu)\delta^{-\frac{1}{2}}]}{2\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}^2} \\
&+ \frac{(a-\mu)\delta^{-2}\phi'[(a-\mu)\delta^{-\frac{1}{2}}]+\delta^{-\frac{3}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{2\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}} \\
&= \frac{(a-\mu)\delta^{-2}\phi'[(a-\mu)\delta^{-\frac{1}{2}}]+3\delta^{-\frac{3}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{2\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}}, \\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\delta^2}\right] &= \frac{1}{2\delta^2}-\delta^{-2}-\frac{(a-\mu)\delta^{-\frac{5}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]} \\
&+ \frac{(a-\mu)^2\delta^{-3}\phi^2[(a-\mu)\delta^{-\frac{1}{2}}]}{4\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}^2} \\
&+ \frac{(a-\mu)^2\delta^{-3}\phi'[(a-\mu)\delta^{-\frac{1}{2}}]+3(a-\mu)\delta^{-\frac{5}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{4\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}} \\
&= -\frac{1}{2\delta^2}+\frac{(a-\mu)^2\delta^{-3}\phi^2[(a-\mu)\delta^{-\frac{1}{2}}]}{4\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}^2} \\
&+ \frac{(a-\mu)^2\delta^{-3}\phi'[(a-\mu)\delta^{-\frac{1}{2}}]-(a-\mu)\delta^{-\frac{5}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{4\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}},
\end{aligned}
$$

$$
E\left[\frac{1}{n}\mathrm{d}^2Q_{\mathrm{MM}}(\theta^\infty|\theta^\infty)\right]=\begin{pmatrix} \frac{1}{n}E\left[\frac{\partial^2Q_{\mathrm{MM}}}{\partial\mu^2}\right] & \frac{1}{n}E\left[\frac{\partial^2Q_{\mathrm{MM}}}{\partial\mu\partial\delta}\right] \\ \frac{1}{n}E\left[\frac{\partial^2Q_{\mathrm{MM}}}{\partial\mu\partial\delta}\right] & \frac{1}{n}E\left[\frac{\partial^2Q_{\mathrm{MM}}}{\partial\delta^2}\right] \end{pmatrix},
$$

$$
\begin{aligned}
\frac{1}{n}E\left[\frac{\partial^2Q_{\mathrm{MM}}}{\partial\mu^2}\right] &= -\frac{1+s_1}{\delta}, \\
\frac{1}{n}E\left[\frac{\partial^2Q_{\mathrm{MM}}}{\partial\mu\partial\delta}\right] &= \frac{s_1g(a;\mu,\delta,-\infty,a)}{\delta}-\frac{\delta^{\frac{3}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]},
\end{aligned}
$$

$$\begin{aligned}\frac{1}{n}E\left[\frac{\partial^2 Q_{\rm MM}}{\partial \delta^2}\right] &= -\frac{1+s_1}{2\delta^2} + \frac{s_1(a-\mu)g(a;\mu,\delta,-\infty,a)}{\delta^2} \\ &\quad - \frac{(a-\mu)\delta^{-\frac{5}{2}}\phi[(a-\mu)\delta^{-\frac{1}{2}}]}{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]},\end{aligned}$$

$$E\left[\frac{1}{n}\mathrm{d}^2 Q_{\rm EM}(\theta^\infty|\theta^\infty)\right] = \begin{pmatrix} \frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial\mu^2}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial\mu\partial\delta}\right] \\ \frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial\mu\partial\delta}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial\delta^2}\right] \end{pmatrix},$$

$$\begin{aligned}\frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial \mu^2}\right] &= -\frac{1}{\delta}, \\ \frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial \mu\partial\delta}\right] &= 0, \\ \frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial \delta^2}\right] &= -\frac{1}{2\delta^2} - \frac{\mu^2\{1-\Phi[(a-\mu)\delta^{-\frac{1}{2}}]\}}{\delta^3}.\end{aligned}$$

Multivariate compound zero-inflated generalized Poisson distribution: First provide some notations below

$$\gamma_1 = \phi_0 + (1-\phi_0)\prod_{i=1}^m[\phi_i + (1-\phi_i)e^{-\lambda_i}],$$

$$a_0 = \prod_{i=1}^m[\phi_i + (1-\phi_i)e^{-\lambda_i}],$$

$$a_i = \prod_{k\neq i}^m[\phi_k + (1-\phi_k)e^{-\lambda_k}],$$

$$a_{il} = \prod_{k\neq i,l}^m[\phi_k + (1-\phi_k)e^{-\lambda_k}],$$

$$\psi_i = \frac{(1-\phi_i)e^{-\lambda_i}}{\phi_i + (1-\phi_i)e^{-\lambda_i}},$$

$$\tau_i = (1-\phi_i)(1-e^{-\lambda_i})(1-\gamma_1),$$

$$\eta_i = \gamma_1 - \gamma_1\psi_i + \psi_i\phi_0 + \phi_i(1-\gamma_1) + \tau_i.$$

The rate matrices of CZIGP via MM and EM algorithms are given by

$$
\begin{aligned}
E[dM_{\mathrm{MM}}(\boldsymbol{\theta}^{\infty})] &= \mathbf{I} - E\left[\frac{d^2Q_{\mathrm{MM}}(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty})}{n}\right]^{-1} E\left[\frac{d^2\ell(\boldsymbol{\theta}^{\infty})}{n}\right], \\
E[dM_{\mathrm{EM}}(\boldsymbol{\theta}^{\infty})] &= \mathbf{I} - E\left[\frac{d^2Q_{\mathrm{EM}}(\boldsymbol{\theta}^{\infty}|\boldsymbol{\theta}^{\infty})}{n}\right]^{-1} E\left[\frac{d^2\ell(\boldsymbol{\theta}^{\infty})}{n}\right],
\end{aligned}
\tag{S3.8}
$$

where

$$
E\left[\frac{1}{n}d^2\ell(\theta^{\infty})\right] = \begin{pmatrix}
\frac{1}{n}E[\frac{\partial^2\ell}{\partial\phi_0^2}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\phi_0\partial\boldsymbol{\phi}^{\top}}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\phi_0\partial\boldsymbol{\lambda}^{\top}}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\phi_0\partial\boldsymbol{\pi}^{\top}}] \\
\frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\phi}\partial\phi_0}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^{\top}}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\phi}\partial\boldsymbol{\lambda}^{\top}}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\phi}\partial\boldsymbol{\pi}^{\top}}] \\
\frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\lambda}\partial\phi_0}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\phi}^{\top}}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^{\top}}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\pi}^{\top}}] \\
\frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\pi}\partial\phi_0}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\pi}\partial\boldsymbol{\phi}^{\top}}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\pi}\partial\boldsymbol{\lambda}^{\top}}] & \frac{1}{n}E[\frac{\partial^2\ell}{\partial\boldsymbol{\pi}\partial\boldsymbol{\pi}^{\top}}]
\end{pmatrix},
$$

$$
\begin{aligned}
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_0^2}\right] &= -\frac{(1-a_0)^2}{\gamma_1} - \frac{1-a_0}{1-\phi_0}, \\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_i^2}\right] &= -\frac{(1-\phi_0)^2(1-e^{-\lambda_i})^2a_i^2}{\gamma_1} - \frac{(1-\gamma_1)(1-e^{-\lambda_i})^2}{\phi_i+(1-\phi_i)e^{-\lambda_i}} \\
&\quad - \frac{(1-e^{-\lambda_i})(1-\gamma_1)}{1-\phi_i}, \\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\lambda_i^2}\right] &= (1-\phi_0)(1-\phi_i)e^{-\lambda_i}a_i - \frac{(1-\phi_0)^2(1-\phi_i)^2e^{-\lambda_i}a_i^2}{\gamma_1} \\
&\quad + \tau_i\left(\frac{\pi_i}{\lambda_i+2\pi_i} - \frac{1}{\lambda_i}\right) + \frac{\phi_i(1-\phi_i)e^{-\lambda_i}(1-\gamma_1)}{\phi_i+(1-\phi_i)e^{-\lambda_i}}, \\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\pi_i^2}\right] &= -\tau_i\left(\frac{2\lambda_i}{\lambda_i+2\pi_i} + \frac{\lambda_i}{1-\pi_i}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_0\partial\phi_i}\right] &= -\frac{(1-e^{-\lambda_i})a_i}{\gamma_1},\\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_0\partial\lambda_i}\right] &= (1-\phi_i)e^{-\lambda_i}a_i + \frac{(1-\phi_0)(1-\phi_i)e^{-\lambda_i}(1-a_0)a_i}{\gamma_1},\\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_i\partial\phi_{l\neq i}}\right] &= (1-\phi_0)(1-e^{-\lambda_i})(1-e^{-\lambda_l})a_{il}\\
&\quad - \frac{(1-\phi_0)^2(1-e^{-\lambda_i})(1-e^{-\lambda_l})a_ia_l}{\gamma_1},\\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_i\partial\lambda_i}\right] &= (1-\phi_0)e^{-\lambda_i}a_i + \frac{(1-\phi_0)^2(1-e^{-\lambda_i})(1-\phi_i)e^{-\lambda_i}a_i^2}{\gamma_1}\\
&\quad + \frac{(1-\gamma_1)e^{-\lambda_i}}{\phi_i+(1-\phi_i)e^{-\lambda_i}},\\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_i\partial\lambda_{l\neq i}}\right] &= \frac{(1-\phi_0)^2(1-e^{-\lambda_i})a_i(1-\phi_l)e^{-\lambda_l}a_l}{\gamma_1}\\
&\quad - (1-\phi_0)(1-e^{-\lambda_i})(1-\phi_l)e^{-\lambda_l}a_{il},\\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\lambda_i\partial\lambda_{l\neq i}}\right] &= (1-\phi_0)(1-\phi_i)(1-\phi_l)e^{-\lambda i}e^{-\lambda_l}a_{il}\\
&\quad - \frac{(1-\phi_0)^2(1-\phi_i)(1-\phi_l)e^{-\lambda_i}e^{-\lambda_l}a_ia_l}{\gamma_1},\\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\lambda_i\partial\pi_i}\right] &= -\frac{\tau_i\lambda_i}{\lambda_i+2\pi_i},\\
\frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\lambda_i\partial\pi_{l\neq i}}\right] &= \frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_0\partial\pi_i}\right] = \frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\phi_i\partial\pi_l}\right] = \frac{1}{n}E\left[\frac{\partial^2\ell}{\partial\pi_i\partial\pi_{l\neq i}}\right] = 0.
\end{aligned}
$$

$$
\begin{aligned}
& E\left[\frac{1}{n}d^2Q_{\mathrm{MM}}(\theta^\infty \mid \theta^\infty)\right]\\
=\ & \mathrm{diag}\bigg(\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial\phi_0^2}\right], \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial\phi_1^2}\right], \dots, \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial\phi_m^2}\right], \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial\lambda_1^2}\right],\\
& \qquad \dots, \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial\lambda_m^2}\right], \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial\pi_1^2}\right], \dots, \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial\pi_m^2}\right]\bigg),
\end{aligned}
$$

$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial \phi_0^2}\right] = -\frac{1}{\phi_0} - \frac{1}{1-\phi_0},$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial \phi_i^2}\right] = -\frac{\gamma_1-\phi_0}{\phi_i[\phi_i+(1-\phi_i)e^{-\lambda i}]} - \frac{(\gamma_1-\phi_0)e^{-\lambda i}}{(1-\phi_i)[\phi_i+(1-\phi_i)e^{-\lambda i}]}$$
$$-\frac{1-\gamma_1}{\phi_i} - \frac{1-\gamma_1}{1-\phi_i},$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial \lambda_i^2}\right] = -\frac{(1-\phi_i)(1-e^{-\lambda i})(1-\gamma_1)}{\lambda_i},$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{MM}}}{\partial \pi_i^2}\right] = -\frac{\tau_i\lambda_i}{\pi_i(1-\pi_i)},\ i=1,\ldots,m.$$

$$E\left[\frac{d^2 Q_{\mathrm{EM}}(\theta^\infty|\theta^\infty)}{n}\right]=\begin{pmatrix} \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\phi_0^2}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\phi_0\partial\boldsymbol{\phi}^\top}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\phi_0\partial\boldsymbol{\lambda}^\top}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\phi_0\partial\boldsymbol{\pi}^\top}\right] \\ \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\phi}\partial\phi_0}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\phi}\partial\boldsymbol{\lambda}^\top}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\phi}\partial\boldsymbol{\pi}^\top}\right] \\ \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\lambda}\partial\phi_0}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\phi}^\top}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\pi}^\top}\right] \\ \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\pi}\partial\phi_0}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\pi}\partial\boldsymbol{\phi}^\top}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\pi}\partial\boldsymbol{\lambda}^\top}\right] & \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial\boldsymbol{\pi}\partial\boldsymbol{\pi}^\top}\right] \end{pmatrix},$$

$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \phi_0^2}\right] = -\frac{1}{\phi_0} - \frac{1}{1-\phi_0},$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \phi_0\partial\phi_i}\right] = \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \phi_0\partial\lambda_i}\right] = \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \phi_0\partial\pi_i}\right] = 0,$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \phi_i^2}\right] = -\frac{1}{\phi_i^2} + \left[\frac{1}{\phi_i^2} - \frac{1}{(1-\phi_i)^2}\right]\cdot[\gamma_1\psi_i + (1-\phi_i-\psi_i)\phi_0$$
$$+(1-\phi_i)\mathrm{e}^{-\lambda_i}(1-\gamma_1)],$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \phi_i\partial\phi_{l\neq i}}\right] = \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \phi_i\partial\lambda_l}\right] = \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \phi_i\partial\pi_l}\right] = 0,$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \lambda_i^2}\right] = -\frac{\eta_i}{\lambda_i} + \frac{\pi_i\eta_i}{\lambda_i+2\pi_i},$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \lambda_i\partial\lambda_{l\neq i}}\right] = \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \lambda_i\partial\pi_{l\neq i}}\right] = \frac{1}{n}E\left[\frac{\partial^2 Q_{\mathrm{EM}}}{\partial \pi_i\partial\pi_{l\neq i}}\right] = 0,$$

$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial\lambda_i\partial\pi_i}\right] = -\frac{\lambda_i\eta_i}{\lambda_i+2\pi_i},$$
$$\frac{1}{n}E\left[\frac{\partial^2 Q_{\rm EM}}{\partial\pi_i^2}\right] = -\frac{\lambda_i\eta_i}{1-\pi_i} - \frac{2\lambda_i\eta_i}{\lambda_i+2\pi_i}.$$

S4 The proof that the supporting hyperplane inequality can be implied by the Jensen's inequality

Statement: **The Jensen's inequality implies the supporting hyperplane inequality.**

Assume that $\psi(\cdot)$ is a convex function, according to the following Jensen's inequality,

$$\psi(\sum_{i=1}^n a_i x_i) \leqslant \sum_{i=1}^n a_i\psi(x_i),$$

where $a_i \geqslant 0$ and $\sum_{i=1}^n a_i = 1$. Simply taking $n=2$, we have $\psi[ax_1+(1-a)x_2] \leqslant a\psi(x_1)+(1-a)\psi(x_2)$, and we can rewrite as

$$\frac{\psi[x_2+a(x_1-x_2)]-\psi(x_2)}{a} \leqslant \psi(x_1)-\psi(x_2),$$

where $a\in(0,1)$. Without loss of generality, let $x_1 \neq x_2$ and let $a\to 0$, we have

$$(x_1-x_2)\lim_{a\to 0}\frac{\psi[x_2+a(x_1-x_2)]-\psi(x_2)}{a(x_1-x_2)} \leqslant \psi(x_1)-\psi(x_2),$$

which is equivalent to $(x_1-x_2)\psi'(x_2) \leqslant \psi(x_1)-\psi(x_2)$.

S5 Some applications of Section 4 in the old version

Generalized Poisson distribution

In this part, we develop an AD-MM algorithm for calculating the MLEs for the *generalized Poisson* (GP) distribution, where the explicit solutions to the MLEs are not available and the EM algorithm does not yet exist due to the absence of latent variables.

A non-negative integer valued random variable Y is said to have the GP distribution with parameters $\lambda > 0$ and π, denoted by $Y \sim \text{GP}(\lambda, \pi)$, if its pmf is given by

$$p(y|\lambda, \pi) = \begin{cases} \dfrac{\lambda(\lambda + \pi y)^{y-1} \mathrm{e}^{-\lambda - \pi y}}{y!}, & y = 0, 1, \ldots, \infty, \\[2ex] 0, & \text{for } y > r, \text{ when } \pi < 0, \end{cases} \tag{S5.9}$$

where $\max(-1, -\lambda/r) < \pi \leqslant 1$ and $r\,(\geqslant 4)$ is the largest positive integer for which $\lambda + \pi r > 0$ when $\pi < 0$. The $\text{GP}(\lambda, \pi)$ distribution reduces to the usual $\text{Poisson}(\lambda)$ when $\pi = 0$, and it has the twin properties of over-dispersion when $\pi > 0$ and under-dispersion when $\pi < 0$. The most frequently used version of the GP distribution assumes $\lambda > 0$ and $\pi \in [0, 1)$.

Let $Y_1, \ldots, Y_n \stackrel{\text{iid}}{\sim} \text{GP}(\lambda, \pi)$ and $Y_{\text{obs}} = \{y_i\}_{i=1}^n$ denote the observed counts. Let $\mathbb{I}_0 = \{i\colon\, y_i = 0,\ 1 \leqslant i \leqslant n\}$, $\mathbb{I}_1 = \{i\colon\, y_i = 1,\ 1 \leqslant i \leqslant n\}$, $\mathbb{I}_2 = \{i\colon\, y_i \geqslant 2,\ 1 \leqslant i \leqslant n\}$, and m_k denote the number of elements in $\mathbb{I}_k$

for $k = 0, 1, 2$. Clearly, we have $m_0 + m_1 + m_2 = n$. The observed-data likelihood function is given by

$$L(\lambda, \pi | Y_{\rm obs}) = \prod_{i \in \mathbb{I}_0} {\rm e}^{-\lambda} \cdot \prod_{i \in \mathbb{I}_1} \lambda {\rm e}^{-\lambda-\pi} \cdot \prod_{i \in \mathbb{I}_2} \frac{\lambda(\lambda + \pi y_i)^{y_i - 1} {\rm e}^{-\lambda-\pi y_i}}{y_i!}.$$

Since $\sum_{i \in \mathbb{I}_2} y_i = n\bar{y} - m_1$, the log-likelihood function can be decomposed as

$$\begin{aligned} \ell(\lambda, \pi | Y_{\rm obs}) &= c + (m_1 + m_2)\log(\lambda) + n(-\lambda) + n\bar{y}(-\pi) \\ &\quad + \sum_{i \in \mathbb{I}_2} (y_i - 1)\log(\lambda + \pi y_i) \\ &\hat{=} \ell_0(\lambda, \pi) + \sum_{i \in \mathbb{I}_2} \ell_i(\boldsymbol{a}_i^{\top} \boldsymbol{\theta}), \end{aligned} \tag{S5.10}$$

where c is a constant not involving (λ, π);

- $\ell_0(\lambda, \pi) = \ell_0(\lambda) + \ell_0(\pi)$ is completely additively separable, $\ell_0(\lambda) = c + (m_1 + m_2)\log(\lambda) + n(-\lambda) \in \mathsf{LG}(\lambda)$ contains two complemental assemblies $\{\log\lambda, -\lambda\}$, and $\ell_0(\pi) = n\bar{y}(-\pi)$ includes one assembly $-\pi$;

- $\ell_i(\cdot) = (y_i - 1)\log(\cdot)$ is a concave function defined in $\mathbb{R}_+$, $\boldsymbol{a}_i = (1, y_i)^{\top}$, and $\boldsymbol{\theta} = (\lambda, \pi)^{\top}$.

In other words, (S5.10) is a special case of (3.2) with $p_i = 2$, $\boldsymbol{h}_i(\boldsymbol{\theta}) = \boldsymbol{\theta}$ for all i, and $n_2 = 0$. Therefore, from (3.3) and (3.5), we have

$$Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = c_i^{(t)} + (y_i - 1)\left[\frac{\lambda^{(t)}}{\beta_i^{(t)}}\log(\lambda) + \frac{\pi^{(t)}y_i}{\beta_i^{(t)}}\log(\pi)\right],\ \beta_i^{(t)} \mathrel{\hat=} \lambda^{(t)} + \pi^{(t)}y_i,$$

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \ell_0(\lambda,\pi) + \sum_{i\in\mathbb{I}_2} Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = c^* + Q_{[\mathrm{I}]}(\lambda|\boldsymbol{\theta}^{(t)}) + Q_{[\mathrm{II}]}(\pi|\boldsymbol{\theta}^{(t)}),$$

where $\{c_i^{(t)}, c^*\}$ are constants not depending on $\boldsymbol{\theta}$, $Q(\cdot|\boldsymbol{\theta}^{(t)})$ is completely additively separable,

$$\begin{cases} Q_{[\mathrm{I}]}(\lambda|\boldsymbol{\theta}^{(t)}) &= \left(m_1 + m_2 + \lambda^{(t)}\sum_{i\in\mathbb{I}_2}\frac{y_i-1}{\beta_i^{(t)}}\right)\log(\lambda) - n\lambda \\ &= \left(n + \lambda^{(t)}\sum_{i=1}^{n}\frac{y_i-1}{\beta_i^{(t)}}\right)\log(\lambda) - n\lambda \in \mathsf{LG}(\lambda), \\ Q_{[\mathrm{II}]}(\pi|\boldsymbol{\theta}^{(t)}) &= \pi^{(t)}\left[\sum_{i=1}^{n}\frac{(y_i-1)y_i}{\beta_i^{(t)}}\right]\log(\pi) - n\bar{y}\pi \in \mathsf{LG}(\pi). \end{cases}$$

In the derivation of $Q_{[\mathrm{I}]}(\lambda|\boldsymbol{\theta}^{(t)})$, we used the following identity:

$$\begin{aligned} \sum_{i=1}^{n}\frac{y_i-1}{\beta_i^{(t)}} &= \left(\sum_{i\in\mathbb{I}_0} + \sum_{i\in\mathbb{I}_1} + \sum_{i\in\mathbb{I}_2}\right)\frac{y_i-1}{\beta_i^{(t)}} \\ &= \sum_{i\in\mathbb{I}_0}\frac{-1}{\lambda^{(t)}} + 0 + \sum_{i\in\mathbb{I}_2}\frac{y_i-1}{\beta_i^{(t)}} = -\frac{m_0}{\lambda^{(t)}} + \sum_{i\in\mathbb{I}_2}\frac{y_i-1}{\beta_i^{(t)}}. \end{aligned}$$

Therefore, we have the following MM iteration:

$$\lambda^{(t+1)} = \frac{n + \lambda^{(t)}\sum_{i=1}^{n}[(y_i-1)/\beta_i^{(t)}]}{n},\quad \pi^{(t+1)} = \frac{\pi^{(t)}\sum_{i=1}^{n}[(y_i-1)y_i/\beta_i^{(t)}]}{n\bar{y}}. \tag{S5.11}$$

Zero-truncated binomial distribution

A discrete random variable Y is said to follow a *zero-truncated binomial*

(ZTB) distribution, denoted by $Y \sim \text{ZTB}(m, \pi)$, if its pmf is

$$\Pr(Y = y) = \frac{1}{1-(1-\pi)^m} \cdot \binom{m}{y} \pi^y (1-\pi)^{m-y}, \quad y = 1, 2, \ldots, m.$$

Let $Y_1, \ldots, Y_n \stackrel{\text{iid}}{\sim} \text{ZTB}(m, \pi)$, $Y_{\text{obs}} = \{y_i\}_{i=1}^n$ denote the observed data and $\bar{y} = (1/n)\sum_{i=1}^n y_i$. The observed-data log-likelihood function of π is

$$\ell(\pi | Y_{\text{obs}}) = \ell_0(\pi) + \ell_3(\pi), \tag{S5.12}$$

where $\ell_0(\pi) = n\bar{y}\log(\pi) + n(m - \bar{y})\log(1-\pi) \in \mathsf{LB}(\pi)$ and $\ell_3(\pi) = -n\log[1-(1-\pi)^m]$.

The first MM algorithm based on the LB function family

Note that $\ell_0(\pi) \in \mathsf{LB}(\pi)$, which guides us to yield an assembly $\log(\pi)$ or $\log(1-\pi)$ from a minorizing function of $\ell_3(\pi)$. Obviously, $\ell(\pi|Y_{\text{obs}})$ in (S5.12) is a special case of (3.6) with $b_i = n$, $\boldsymbol{a}_i^{\top}\boldsymbol{h}_i(\boldsymbol{\theta}) = (1-\pi)^m$ and $n_3 = 1$. From (3.9), we obtain

$$\begin{aligned} Q(\pi|\pi^{(t)}) &= c + \ell_0(\pi) + \frac{nm(1-\pi^{(t)})^m}{1-(1-\pi^{(t)})^m}\log(1-\pi) \\ &= c + n\bar{y}\log(\pi) + \left[\frac{nm}{1-(1-\pi^{(t)})^m} - n\bar{y}\right]\log(1-\pi) \in \mathsf{LB}(\pi), \end{aligned}$$

minorizing the log-likelihood function $\ell(\pi|Y_{\text{obs}})$. The first MM iteration is

$$\pi^{(t+1)} = \frac{\bar{y}[1-(1-\pi^{(t)})^m]}{m}. \tag{S5.13}$$

The second MM algorithm based on the LEB function family

If we could find an assembly $-\pi$ from a minorizing function of $\ell_3(\pi)$, then the global surrogate function belongs to the LEB function family, resulting in an explicit solution. Let $u = g(\pi) = 1 - \pi$ and $\ell_3(u) = -n\log(1-u^m)$. Since

$$\ell_3'(u) = \frac{nmu^{m-1}}{1-u^m} > 0 \quad \text{and} \quad \ell_3''(u) = \frac{nm(m-1)u^{m-2}}{1-u^m} + \frac{nm^2u^{2m-2}}{(1-u^m)^2} > 0,$$

$\ell_3(u)$ is strictly convex. By applying (3.4), we obtain

$$Q_2(\pi|\pi^{(t)}) = c_2 - \frac{nm(1-\pi^{(t)})^{m-1}}{1-(1-\pi^{(t)})^m} \cdot \pi,$$

minorizing $\ell_3(\pi)$ so that

$$Q(\pi|\pi^{(t)}) = c_2 + \ell_0(\pi) + \frac{nm(1-\pi^{(t)})^{m-1}}{1-(1-\pi^{(t)})^m} \cdot (-\pi) \in \mathsf{LEB}(\pi),$$

minorizing the log-likelihood function $\ell(\pi|Y_{\text{obs}})$. The second MM iteration is

$$\pi^{(t+1)} = \frac{a^{(t)} + b^{(t)} - \sqrt{(a^{(t)}+b^{(t)})^2 - 4\bar{y}a^{(t)}b^{(t)}/m}}{2a^{(t)}}. \tag{S5.14}$$

where $a^{(t)} = (1-\pi^{(t)})^{m-1}$ and $b^{(t)} = 1-(1-\pi^{(t)})^m$.

Multivariate Poisson distribution

Let $X_i = W_0 + W_i$ for $i = 1, \ldots, m$ and $\{W_i\}_{i=0}^m \overset{\text{ind}}{\sim} \text{Poisson}\,(\lambda_i)$. Then, the discrete random vector $\mathbf{x} = (X_1, \ldots, X_m)^\top$ is said to follow an m-dimensional Poisson distribution with parameters $\lambda_0 \geqslant 0$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}_+^m$, denoted by $\mathbf{x} \sim \text{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$ or $\mathbf{x} \sim \text{MP}_m(\lambda_0, \boldsymbol{\lambda})$, accordingly. The joint pmf of $\mathbf{x}$ is

$$\Pr(\mathbf{x}=\boldsymbol{x}) = \sum_{k=0}^{\min(\boldsymbol{x})} \frac{\lambda_0^k \mathrm{e}^{-\lambda_0}}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{x_i-k}\mathrm{e}^{-\lambda_i}}{(x_i-k)!},$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$, $\{x_i\}_{i=1}^m$ are the corresponding realizations of $\{X_i\}_{i=1}^m$, and $\min(\boldsymbol{x}) \,\hat{=}\, \min(x_1, \ldots, x_m)$.

Let $\{\mathbf{x}_j\}_{j=1}^n \overset{\text{iid}}{\sim} \mathrm{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$ and $Y_{\text{obs}} = \{\boldsymbol{x}_j\}_{j=1}^n$ denote the observed data, where $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{mj})^\top$ is the realization of $\mathbf{x}_j = (X_{1j}, \ldots, X_{mj})^\top$. For convenience's sake, we define $\boldsymbol{\theta} = (\lambda_0, \lambda_1, \ldots, \lambda_m)^\top$, $p_j = \min(\boldsymbol{x}_j)$,

$$\begin{aligned} h_{jk}(\boldsymbol{\theta}) &= \frac{\lambda_0^k \mathrm{e}^{-\lambda_0}}{k!} \cdot \frac{\lambda_1^{x_{1j}-k}\mathrm{e}^{-\lambda_1}}{(x_{1j}-k)!} \cdots \frac{\lambda_m^{x_{mj}-k}\mathrm{e}^{-\lambda_m}}{(x_{mj}-k)!}, \quad \text{and} \\ b_{jk}^{(t)} &= \frac{h_{jk}(\boldsymbol{\theta}^{(t)})}{\mathbf{1}_{p_j+1}^\top \boldsymbol{h}_j(\boldsymbol{\theta}^{(t)})}, \qquad j=1,\ldots,n, \quad k=0,1,\ldots,p_j. \end{aligned}$$

The observed-data log-likelihood function of $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = \sum_{j=1}^n \log\left[h_{j0}(\boldsymbol{\theta}) + h_{j1}(\boldsymbol{\theta}) + \cdots + h_{jp_j}(\boldsymbol{\theta})\right] = \sum_{j=1}^n \log\left[\mathbf{1}_{p_j+1}^\top \boldsymbol{h}_j(\boldsymbol{\theta})\right], \tag{S5.15}$$

where $\boldsymbol{h}_j(\boldsymbol{\theta}) = [h_{j0}(\boldsymbol{\theta}), h_{j1}(\boldsymbol{\theta}), \ldots, h_{jp_j}(\boldsymbol{\theta})]^\top$. Clearly, (S5.15) is a special case of (3.2) with $\ell_0(\boldsymbol{\theta}) = 0$ and $n_2 = 0$. Hence, from (3.3) and (3.5), we have

$$
\begin{aligned}
Q_j(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \sum_{k=0}^{p_j} \frac{h_{jk}(\boldsymbol{\theta}^{(t)})}{\mathbf{1}^\top \boldsymbol{h}_j(\boldsymbol{\theta}^{(t)})} \log\left[\frac{\mathbf{1}^\top \boldsymbol{h}_j(\boldsymbol{\theta}^{(t)})}{h_{jk}(\boldsymbol{\theta}^{(t)})} \cdot h_{jk}(\boldsymbol{\theta})\right] \\
&= c_j^{(t)} + \left(\sum_{k=0}^{p_j} k b_{jk}^{(t)}\right)\log(\lambda_0) - \lambda_0 \\
&\quad + \sum_{i=1}^{m}\left\{\left[\sum_{k=0}^{p_j}(x_{ij}-k)b_{jk}^{(t)}\right]\log(\lambda_i) - \lambda_i\right\}, \\
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \sum_{j=1}^{n} Q_j(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \sum_{j=1}^{n} c_j^{(t)} + Q_{[0]}(\lambda_0|\boldsymbol{\theta}^{(t)}) + \sum_{i=1}^{m} Q_{[i]}(\lambda_i|\boldsymbol{\theta}^{(t)}),
\end{aligned}
$$

where $\{c_j^{(t)}\}$ are constants not involving $\boldsymbol{\theta}$, $Q(\cdot|\boldsymbol{\theta}^{(t)})$ is completely additively separable, (3.2) with $\ell_0(\boldsymbol{\theta}) = 0$ and $n_2 = 0$. Hence, from (3.3) and (3.5), we have

$$
\begin{aligned}
Q_{[0]}(\lambda_0|\boldsymbol{\theta}^{(t)}) &= \left(\sum_{j=1}^{n}\sum_{k=0}^{p_j} k b_{jk}^{(t)}\right)\log(\lambda_0) - n\lambda_0 \in \mathsf{LG}(\lambda_0), \\
Q_{[i]}(\lambda_i|\boldsymbol{\theta}^{(t)}) &= \left[\sum_{j=1}^{n}\sum_{k=0}^{p_j}(x_{ij}-k) b_{jk}^{(t)}\right]\log(\lambda_i) - n\lambda_i \in \mathsf{LG}(\lambda_i).
\end{aligned}
$$

Therefore, the MM iterations are given by

$$
\lambda_0^{(t+1)} = \frac{\sum_{j=1}^{n}\sum_{k=0}^{p_j} k b_{jk}^{(t)}}{n}, \quad \lambda_i^{(t+1)} = \frac{\sum_{j=1}^{n}\sum_{k=0}^{p_j}(x_{ij}-k) b_{jk}^{(t)}}{n}, \tag{S5.16}
$$

for $i = 1, \ldots, m$.

S6 Some simulation results of Section 5 in the old version

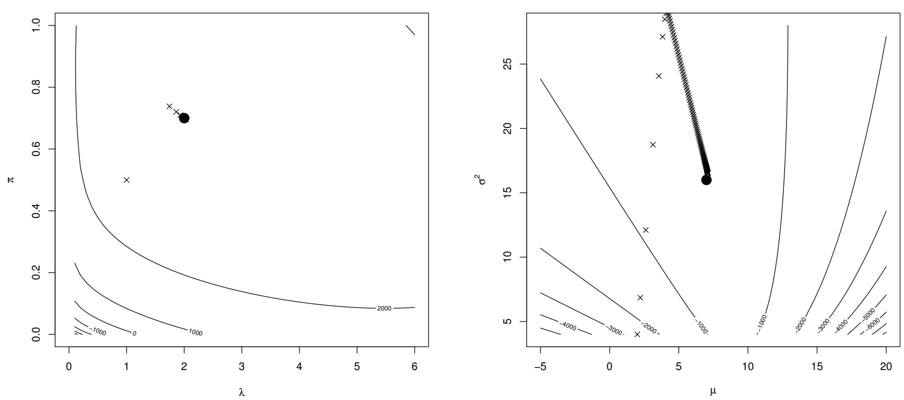

Figure 1: The MM iteration points (marked with "×") on the contour plots of the log-likelihood functions for the generalized Poisson (GP) distribution (the left one) and for the left-truncated normal (LTN) distribution (the right one) converged to their stationary points marked with "•", respectively.

S7 Section 6 in the old version: Further extensions

When the objective function is of the form of $\{g_l(\boldsymbol{a}^\top\boldsymbol{\theta})\}_{l=1}^7$

In Subsection 3.1, we have introduced seven one-dimensional functions $\{g_l(\theta)\}_{l=1}^7$ and the q-dimensional function $g_8(\boldsymbol{\theta})$, where each function $g_l(\theta)$ is a linear combination of some assemblies and complemental assemblies. In this subsection, we extend $\{g_l(\theta)\}_{l=1}^7$ to $\{g_l(\boldsymbol{a}^\top\boldsymbol{\theta})\}_{l=1}^7$ by replacing θ with $\boldsymbol{a}^\top\boldsymbol{\theta}$ since $\boldsymbol{a}^\top\boldsymbol{\theta}$ is usually appeared in regression models. Let $\boldsymbol{a} = (a_1, \ldots, a_q)^\top$

and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$. We assume that

$$a_j \geqslant 0, \quad \theta_j \geqslant 0 \quad \text{for} \quad j = 1, \ldots, q. \tag{S7.17}$$

When the conditions in (S7.17) are violated, we will give a discussion at the end of Subsection 7.2. Under (S7.17), seven separable functions $\{Q_l(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})\}_{l=1}^7$ can be constructed such that

$$Q_l(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \leqslant g_l(\boldsymbol{a}^\top\boldsymbol{\theta}), \quad \forall\, \boldsymbol{\theta}, \boldsymbol{\theta}^{(t)} \in \boldsymbol{\Theta} \quad \text{and} \quad Q_l(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}) = g_l(\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}).$$

In other words, when the log-likelihood function $\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = g_l(\boldsymbol{a}^\top\boldsymbol{\theta})$ for $l = 1, \ldots, 7$, we can obtain explicit MM iterations for calculating the MLE of $\boldsymbol{\theta}$ based on the proposed AD technique in Section 3.

1) Let $\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = g_1(\boldsymbol{a}^\top\boldsymbol{\theta}) = c_0 + c_1 \log(\boldsymbol{a}^\top\boldsymbol{\theta}) - c_2(\boldsymbol{a}^\top\boldsymbol{\theta})^k$. The goal is to find the MLEs of $\boldsymbol{\theta}$ via an MM algorithm. Since both $\log(x)$ and $-x^k$ ($k \in \{1, 2, \ldots, \infty\}$) are concave for $x > 0$, according to the discrete version of Jensen's inequality (2.3), we obtain

$$\begin{aligned} g_1(\boldsymbol{a}^\top\boldsymbol{\theta}) \quad &\geqslant \quad c_0 + \sum_{j=1}^q \left[c_1\omega_j^{(t)} \log\left(\frac{a_j\theta_j}{\omega_j^{(t)}}\right) - c_2\omega_j^{(t)} \left(\frac{a_j\theta_j}{\omega_j^{(t)}}\right)^k \right] \\ &\hat{=} \quad c_0 + \sum_{j=1}^q G_{1j}\left(\frac{a_j\theta_j}{\omega_j^{(t)}}\bigg|\boldsymbol{\theta}^{(t)}\right) \ \hat{=}\ Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}), \end{aligned}$$

where

$$\omega_j^{(t)} = \frac{a_j\theta_j^{(t)}}{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}} \geqslant 0, \qquad j = 1, \ldots, q, \tag{S7.18}$$

and

$$G_{1j}(\phi_j|\boldsymbol{\theta}^{(t)}) = c_1\omega_j^{(t)}\log(\phi_j) + c_2\omega_j^{(t)}(-\phi_j^k) \in \mathsf{LGG}_k(\phi_j).$$

In other words, at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$, $g_1(\boldsymbol{a}^\top\boldsymbol{\theta})$ majorizes $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$, which is a sum of q separable log-generalized-gamma functions. The explicit MM iterations for calculating the MLEs of $\{\theta_j\}_{j=1}^q$ are given by

$$\theta_j^{(t+1)} = \left(\frac{c_1}{kc_2}\right)^{1/k}\frac{\omega_j^{(t)}}{a_j} = \left(\frac{c_1}{kc_2}\right)^{1/k}\frac{\theta_j^{(t)}}{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}}, \qquad j = 1, \ldots, q,$$

or in the form of vectors

$$\boldsymbol{\theta}^{(t+1)} = \left(\frac{c_1}{kc_2}\right)^{1/k}\frac{\boldsymbol{\theta}^{(t)}}{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}}. \tag{S7.19}$$

2) Let $\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = g_2(\boldsymbol{a}^\top\boldsymbol{\theta}) = c_0 + c_1\log(\boldsymbol{a}^\top\boldsymbol{\theta}) + c_2\log(1 - \boldsymbol{a}^\top\boldsymbol{\theta})$. The goal is to find the MLEs of $\boldsymbol{\theta}$ via an MM algorithm. Since both $\log(x)$ and $\log(1-x)$ are concave for $x \in (0,1)$, we can obtain the following separable minorizing function

$$Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = c_0 + \sum_{j=1}^q G_{2j}\left(\frac{a_j\theta_j}{\omega_j^{(t)}}\middle|\boldsymbol{\theta}^{(t)}\right),$$

where $\{\omega_j^{(t)}\}$ are defined by (S7.18) and

$$G_{2j}(\phi_j|\boldsymbol{\theta}^{(t)}) = c_1\omega_j^{(t)}\log(\phi_j) + c_2\omega_j^{(t)}\log(1-\phi_j) \in \mathsf{LB}(\phi_j).$$

The explicit MM iteration for calculating the MLE of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}^{(t+1)} = \frac{c_1}{c_1 + c_2}\cdot\frac{\boldsymbol{\theta}^{(t)}}{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}}. \tag{S7.20}$$

3) Let $\ell(\boldsymbol{\theta}|Y_{\rm obs}) = g_3(\boldsymbol{a}^\top\boldsymbol{\theta}) = c_0 + c_1\log(\boldsymbol{a}^\top\boldsymbol{\theta}) + c_2\log(1-\boldsymbol{a}^\top\boldsymbol{\theta}) - c_3(\boldsymbol{a}^\top\boldsymbol{\theta})$. Since both $\log(x)$ and $\log(1-x)$ are concave for $x \in (0,1)$, we can obtain the following separable minorizing function

$$Q_3(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = c_0 + \sum_{j=1}^{q} G_{3j}\left(\frac{a_j\theta_j}{\omega_j^{(t)}}\middle|\boldsymbol{\theta}^{(t)}\right),$$

where $\{\omega_j^{(t)}\}$ are defined by (S7.18) and

$$G_{3j}(\phi_j|\boldsymbol{\theta}^{(t)}) = c_1\omega_j^{(t)}\log(\phi_j) + c_2\omega_j^{(t)}\log(1-\phi_j) - c_3\omega_j^{(t)}\phi_j \in \mathsf{LEB}(\phi_j).$$

The explicit MM iteration for calculating the MLE of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}^{(t+1)} = \frac{c_1+c_2+c_3-\sqrt{(c_1+c_2+c_3)^2-4c_1c_3}}{2c_3}\cdot\frac{\boldsymbol{\theta}^{(t)}}{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}}. \qquad \text{(S7.21)}$$

4) Let $\ell(\boldsymbol{\theta}|Y_{\rm obs}) = g_4(\boldsymbol{a}^\top\boldsymbol{\theta}) = c_0 + c_1\log(\boldsymbol{a}^\top\boldsymbol{\theta}) - (c_1+c_2)\log(1+\boldsymbol{a}^\top\boldsymbol{\theta})$. Since $\log(x)$ is concave for $x>0$ and $-\log(1+x)$ is convex for $x>0$, from (2.3) and the supporting hyperplane inequality (2.4), we can obtain the following separable minorizing function

$$Q_4(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = c_0^* + \sum_{j=1}^{q} G_{4j}\left(\frac{a_j\theta_j}{\omega_j^{(t)}}\middle|\boldsymbol{\theta}^{(t)}\right),$$

where $\{\omega_j^{(t)}\}$ are defined by (S7.18) and

$$G_{4j}(\phi_j|\boldsymbol{\theta}^{(t)}) = c_1\omega_j^{(t)}\log(\phi_j) - \frac{(c_1+c_2)\omega_j^{(t)}}{1+\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}}\cdot\phi_j \in \mathsf{LG}(\phi_j).$$

Note that $\mathsf{LG}(\cdot) = \mathsf{LGG}_1(\cdot)$ and we can immediately obtain the MM iteration

$$\boldsymbol{\theta}^{(t+1)} = \frac{c_1(1+\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)})}{c_1+c_2}\cdot\frac{\boldsymbol{\theta}^{(t)}}{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}}. \qquad \text{(S7.22)}$$

5) Let $\ell(\boldsymbol{\theta}|Y_{\rm obs}) = g_5(\boldsymbol{a}^\top\boldsymbol{\theta}) = c_0 + c_1\log(\boldsymbol{a}^\top\boldsymbol{\theta}) - c_2(\boldsymbol{a}^\top\boldsymbol{\theta}) + c_3\log(1+\boldsymbol{a}^\top\boldsymbol{\theta})$. Since both $\log(x)$ and $\log(1+x)$ are concave for $x>0$, we obtain the following separable minorizing function

$$Q_5(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = c_0 + \sum_{j=1}^q G_{5j}\left(\frac{a_j\theta_j}{\omega_j^{(t)}}\bigg|\boldsymbol{\theta}^{(t)}\right),$$

where $\{\omega_j^{(t)}\}$ are defined by (S7.18) and

$$G_{5j}(\phi_j|\boldsymbol{\theta}^{(t)}) = c_1\omega_j^{(t)}\log(\phi_j) - c_2\omega_j^{(t)}\cdot\phi_j + c_3\omega_j^{(t)}\log(1+\phi_j) \in \mathsf{LEG}(\phi_j).$$

The explicit MM iteration for calculating the MLE of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}^{(t+1)} = \frac{c_1 - c_2 + c_3 + \sqrt{(c_1-c_2+c_3)^2 + 4c_1c_2}}{2c_2}\cdot\frac{\boldsymbol{\theta}^{(t)}}{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}}. \tag{S7.23}$$

6) Let $\ell(\boldsymbol{\theta}|Y_{\rm obs}) = g_6(\boldsymbol{a}^\top\boldsymbol{\theta}) = c_0 - c_1\log(\boldsymbol{a}^\top\boldsymbol{\theta}) - c_2(\boldsymbol{a}^\top\boldsymbol{\theta})^{-1}$. Since $-\log(x)$ is convex for $x>0$ and $-1/x$ is concave for $x>0$, we obtain the following separable minorizing function

$$Q_6(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = c_0^* + \sum_{j=1}^q\left(-\frac{c_1a_j\theta_j}{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}} - \frac{c_2\omega_j^{(t)2}}{a_j\theta_j}\right)$$

with the following MM iteration

$$\boldsymbol{\theta}^{(t+1)} = \sqrt{\frac{c_2}{c_1}}\cdot\frac{\boldsymbol{\theta}^{(t)}}{\sqrt{\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}}}. \tag{S7.24}$$

7) Let $\ell(\boldsymbol{\theta}|Y_{\rm obs}) = g_7(\boldsymbol{a}^\top\boldsymbol{\theta}) = c_0 - c_1\exp(-c_2\boldsymbol{a}^\top\boldsymbol{\theta}) - c_3(\boldsymbol{a}^\top\boldsymbol{\theta})$. Since $-\mathrm{e}^x$ is concave for $x\in\mathbb{R}$, we obtain the following separable minorizing function

$$Q_7(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = c_0 + \sum_{j=1}^q G_{7j}\left(\frac{a_j\theta_j}{\omega_j^{(t)}}\bigg|\boldsymbol{\theta}^{(t)}\right),$$

where $\{\omega_j^{(t)}\}$ are defined by (S7.18) and

$$G_{7j}(\phi_j|\boldsymbol{\theta}^{(t)}) = -c_1\omega_j^{(t)}\exp(-c_2\phi_j) - c_3\omega_j^{(t)}\cdot\phi_j \in \mathsf{LGM}(\phi_j).$$

The explicit MM iteration for calculating the MLE of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}^{(t+1)} = \frac{\log(c_1c_2/c_3)}{c_2}\cdot\frac{\boldsymbol{\theta}^{(t)}}{\boldsymbol{a}^{\top}\boldsymbol{\theta}^{(t)}}. \tag{S7.25}$$

When the log-likelihood function is beyond $\{g_l(\boldsymbol{a}^{\top}\boldsymbol{\theta})\}_{l=1}^{7}$

When the log-likelihood function $\ell(\boldsymbol{\theta}|Y_{\text{obs}})$ is beyond $\{g_l(\boldsymbol{a}^{\top}\boldsymbol{\theta})\}_{l=1}^{7}$, we may try to construct a separable surrogate function, $Q^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ say, satisfying that

(i) $Q^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ minorizes $\ell(\boldsymbol{\theta}|Y_{\text{obs}})$; and

(ii) $Q^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ belongs to $\{g_l(\boldsymbol{a}^{\top}\boldsymbol{\theta})\}_{l=1}^{7}$.

In this way, we can obtain explicit MM iterations for calculating the MLE of $\boldsymbol{\theta}$ based on the proposed AD technique in Section 3.

Example 1. We revisit $\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = g_4(\boldsymbol{a}^{\top}\boldsymbol{\theta}) = c_0 + c_1\log(\boldsymbol{a}^{\top}\boldsymbol{\theta}) - (c_1 + c_2)\log(1+\boldsymbol{a}^{\top}\boldsymbol{\theta})$ in Part 4) of Subsection 6.1. Alternatively, we could construct a separable surrogate function $Q_4^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ to minorize $g_4(\boldsymbol{a}^{\top}\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$, where $Q_4^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ is a special case of $g_1(\boldsymbol{a}^{\top}\boldsymbol{\theta})$. In fact, since $-\log(1+x)$ is convex for $x>0$, from the supporting hyperplane inequality

(2.4), we have

$$-\log(1+\boldsymbol{a}^{\top}\boldsymbol{\theta}) \geqslant -\log(1+\boldsymbol{a}^{\top}\boldsymbol{\theta}^{(t)}) - \frac{\boldsymbol{a}^{\top}\boldsymbol{\theta} - \boldsymbol{a}^{\top}\boldsymbol{\theta}^{(t)}}{1+\boldsymbol{a}^{\top}\boldsymbol{\theta}^{(t)}},$$

so that

$$g_4(\boldsymbol{a}^{\top}\boldsymbol{\theta}) \geqslant c_3 + c_1\log(\boldsymbol{a}^{\top}\boldsymbol{\theta}) - \frac{c_1+c_2}{1+\boldsymbol{a}^{\top}\boldsymbol{\theta}^{(t)}}\cdot \boldsymbol{a}^{\top}\boldsymbol{\theta} \;\hat{=}\; Q_4^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}),$$

which is a special case of $g_1(\boldsymbol{a}^{\top}\boldsymbol{\theta})$ with $k=1$. From (S7.19), we immediately obtain the MM iteration

$$\boldsymbol{\theta}^{(t+1)} = \frac{c_1(1+\boldsymbol{a}^{\top}\boldsymbol{\theta}^{(t)})}{c_1+c_2}\cdot\frac{\boldsymbol{\theta}^{(t)}}{\boldsymbol{a}^{\top}\boldsymbol{\theta}^{(t)}}, \tag{S7.26}$$

which is identical to (S7.22).

Example 2. Assume that the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}|Y_{\mathrm{obs}}) = c_0 + \sum_{i=1}^{n}\left\{-\boldsymbol{a}_i^{\top}\boldsymbol{\theta} - \log\left[\sum_{j=1}^{n} c_{ij}\exp(-\boldsymbol{a}_j^{\top}\boldsymbol{\theta})\right]\right\}, \qquad c_{ij}\geqslant 0.$$

Since $-\log(x)$ is convex for $x>0$, from the supporting hyperplane inequality (2.4), we have

$$\begin{aligned}-\log\left[\sum_{j=1}^{n} c_{ij}\exp(-\boldsymbol{a}_j^{\top}\boldsymbol{\theta})\right] \;\geqslant\; & -\log\left[\sum_{j=1}^{n} c_{ij}\exp(-\boldsymbol{a}_j^{\top}\boldsymbol{\theta}^{(t)})\right] \\ & - \frac{\sum_{j=1}^{n} c_{ij}\left[\exp(-\boldsymbol{a}_j^{\top}\boldsymbol{\theta}) - \exp(-\boldsymbol{a}_j^{\top}\boldsymbol{\theta}^{(t)})\right]}{\sum_{j=1}^{n} c_{ij}\exp(-\boldsymbol{a}_j^{\top}\boldsymbol{\theta}^{(t)})}\end{aligned}$$

so that

$$\ell(\boldsymbol{\theta}|Y_{\mathrm{obs}}) \geqslant c_1 + \sum_{i=1}^{n}\left\{-\boldsymbol{a}_i^{\top}\boldsymbol{\theta} - \frac{\sum_{j=1}^{n} c_{ij}\exp(-\boldsymbol{a}_j^{\top}\boldsymbol{\theta})}{\sum_{k=1}^{n} c_{ik}\exp(-\boldsymbol{a}_k^{\top}\boldsymbol{\theta}^{(t)})}\right\} \;\hat{=}\; Q^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}),$$

which minorizes $\ell(\boldsymbol{\theta}|Y_{\text{obs}})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$. Note that

$$\begin{aligned} Q^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= -\sum_{j=1}^{n} \boldsymbol{a}_j^\top \boldsymbol{\theta} - \sum_{j=1}^{n} (\textstyle\sum_{i=1}^n c'_{ij}) \exp(-\boldsymbol{a}_j^\top \boldsymbol{\theta}) \\ &= \sum_{j=1}^{n} \left[-\boldsymbol{a}_j^\top \boldsymbol{\theta} - (\textstyle\sum_{i=1}^n c'_{ij}) \exp(-\boldsymbol{a}_j^\top \boldsymbol{\theta}) \right] = \sum_{j=1}^{n} g_7(\boldsymbol{a}_j^\top \boldsymbol{\theta}), \end{aligned}$$

which is a linear combination of $g_7(\boldsymbol{a}_j^\top \boldsymbol{\theta})$, where

$$c'_{ij} = \frac{c_{ij}}{\sum_{k=1}^n c_{ik} \exp(-\boldsymbol{a}_k^\top \boldsymbol{\theta}^{(t)})}, \quad i, j = 1, \ldots, n.$$

Although we cannot immediately obtain the MM iteration from (S7.25), a similar method can be used to separate the parameters within the vector $\boldsymbol{\theta}$ and to obtain the MM iteration.

S8 Section 7 in the old version: Illustration and summary

In this subsection, we show that most of the inequalities in the MM literature are special cases of Jensen's inequality. Our AD method distinguishes itself from solely using these inequalities in the way that as guided by the A-technique, it decomposes the target function or some intermediate surrogate function into different parts to be minorized separately. Our approach sets a clear goal of constructing a surrogate function as the sum of separable univariate functions for numerical convenience.

(1) The following (S8.27) is the arithmetic-geometric mean inequality, which is used by Lange and Zhou (2014) (p.341) in the unconstrained signomial programming for the terms $c_{\boldsymbol{\alpha}} \prod_{i=1}^n x_i^{\alpha_i}$ with positive coefficients $c_{\boldsymbol{\alpha}}$:

$$\prod_{i=1}^n z_i^{\alpha_i} \leqslant \sum_{i=1}^n \frac{\alpha_i}{\|\boldsymbol{\alpha}\|_1} z_i^{\|\boldsymbol{\alpha}\|_1}, \tag{S8.27}$$

where $\{z_i\}_{i=1}^n$ and $\{\alpha_i\}_{i=1}^n$ are non-negative numbers, and the ℓ_1-norm $\|\boldsymbol{\alpha}\|_1 \hat{=} \sum_{i=1}^n |\alpha_i|$. In fact, the inequality (S8.27) is from Jensen's inequality (2.3) with $\varphi(\cdot) = \log(\cdot)$; that is,

$$\log\left(\sum_{i=1}^n \frac{\alpha_i}{\|\boldsymbol{\alpha}\|_1} z_i^{\|\boldsymbol{\alpha}\|_1}\right) \geqslant \sum_{i=1}^n \frac{\alpha_i}{\|\boldsymbol{\alpha}\|_1} \log\left(z_i^{\|\boldsymbol{\alpha}\|_1}\right) = \sum_{i=1}^n \alpha_i \log(z_i).$$

By taking exponential operation on both sides of the above inequality, we immediately obtain the arithmetic-geometric mean inequality (S8.27).

(2) The following inequality (S8.28) is used by Lange and Zhou (2014) (p.342) in the unconstrained signomial programming for the terms $c_{\boldsymbol{\alpha}} \prod_{i=1}^n x_i^{\alpha_i}$ with $c_{\boldsymbol{\alpha}} < 0$:

$$\prod_{i=1}^n x_i^{\alpha_i} \geqslant \prod_{j=1}^n x_{mj}^{\alpha_j} \left[1 + \sum_{i=1}^n \alpha_i \log(x_i) - \sum_{i=1}^n \alpha_i \log(x_{mi})\right], \tag{S8.28}$$

where $\{x_i\}$ and $\{\alpha_i\}$ are positive numbers, and x_{mi} denotes the m-th approximation of x_i. In fact, the inequality (S8.28) comes from the supporting hyperplane inequality (2.4) with $\psi(z) = -\log(z)$ at the point $z_0 = 1$; that is, $-\log(z) \geqslant -z+1$ or $z \geqslant 1+\log(z)$. In particular, let $z = \prod_{i=1}^n (x_i/x_{mi})^{\alpha_i}$, we immediately obtain the inequality (S8.28).

(3) Let $\mathbb{C}$ be a closed convex set in $\mathbb{R}^k$ and $d(\boldsymbol{x}, \mathbb{C}) \hat{=} \inf\{\|\boldsymbol{x}-\boldsymbol{y}\|\colon \boldsymbol{y} \in \mathbb{C}\}$ denote the Euclidean distance from $\boldsymbol{x}$ in the closed convex set $\mathbb{S} \subset \mathbb{R}^k$ to $\mathbb{C}$. Chi and Lange (2014) (p.98) used the following (S8.29) and (S8.30) to get an MM algorithm for the heron problem:

$$d(\boldsymbol{x}, \mathbb{C}) \leqslant \|\boldsymbol{x} - P_{\mathbb{C}}(\boldsymbol{x}_m)\| \tag{S8.29}$$

$$\leqslant \|\boldsymbol{x}_m - P_{\mathbb{C}}(\boldsymbol{x}_m)\| + \frac{\|\boldsymbol{x}-P_{\mathbb{C}}(\boldsymbol{x}_m)\|^2 - \|\boldsymbol{x}_m - P_{\mathbb{C}}(\boldsymbol{x}_m)\|^2}{2\|\boldsymbol{x}_m - P_{\mathbb{C}}(\boldsymbol{x}_m)\|}, \tag{S8.30}$$

where $\boldsymbol{x}_m$ is the m-th approximation of $\boldsymbol{x}$, $P_{\mathbb{C}}(\boldsymbol{x}_m) \hat{=} \arg\min_{\boldsymbol{y}\in\mathbb{C}} \|\boldsymbol{x}_m - \boldsymbol{y}\|$ denotes the projection of $\boldsymbol{x}_m$ onto the set $\mathbb{C}$. In fact, the inequality (S8.29) follows directly from the definition of the distance function. The inequality (S8.30) comes from the supporting hyperplane inequality (2.4) with $\psi(u) = -\sqrt{u}$ at the point $u_0 = u_m$; that is, $-\sqrt{u} \geqslant -\sqrt{u_m} - (u-u_m)/(2\sqrt{u_m})$. In particular, let $u = \|\boldsymbol{x} - P_{\mathbb{C}}(\boldsymbol{x}_m)\|^2$ and $u_m = \|\boldsymbol{x}_m - P_{\mathbb{C}}(\boldsymbol{x}_m)\|^2$, we immediately obtain the inequality (S8.30).

(4) Let $h_1(\boldsymbol{\beta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2$. The inequality (3.14) in Yen (2011)

can be stated as

$$h_1\Big(\beta^{(l+1)}\Big) + \rho_{\lambda,k,\sigma^2} \lim_{\tau_3\to 0} \rho_{\tau_3} \sum_{j=1}^{p} \log\Big(1+\tau_3^{-1}|\beta_j^{(l+1)}|\Big)$$

$$\leqslant h_1\Big(\beta^{(l+1)}\Big) + \rho_{\lambda,k,\sigma^2} \lim_{\tau_3\to 0} \rho_{\tau_3} \sum_{j=1}^{p} \left[\log\Big(1+\tau_3^{-1}|\beta_j^{(l)}|\Big) + \frac{|\beta_j^{(l+1)}|+\tau_3}{|\beta_j^{(l)}|+\tau_3} - 1\right],$$

which in fact comes from the Taylor expansion of the convex function $-\log(z)$ at the point $z_0 = 1+\tau_3^{-1}|\beta_j^{(l)}|$ with $z = 1+\tau_3^{-1}|\beta_j^{(l+1)}|$.

(5) The inequality (7) in Zhou and Zhang (2012), i.e.,

$$\log(\alpha_j+k) \geqslant \frac{\alpha_j^{(t)}}{\alpha_j^{(t)}+k} \log\left(\frac{\alpha_j^{(t)}+k}{\alpha_j^{(t)}} \cdot \alpha_j\right) + \frac{k}{\alpha_j^{(t)}+k} \log\left(\frac{\alpha_j^{(t)}+k}{k} \cdot k\right)$$

comes from Jensen's inequality on the concave function $\log(x)$.

(6) The inequality (8) in Zhou and Zhang (2012), i.e.,

$$-\log(|\boldsymbol{\alpha}|+k) \geqslant -\log(|\boldsymbol{\alpha}^{(t)}|+k) - \frac{|\boldsymbol{\alpha}| - |\boldsymbol{\alpha}^{(t)}|}{|\boldsymbol{\alpha}^{(t)}|+k}$$

comes from the Taylor expansion of the convex function $-\log(z)$ at the point $z_0 = |\boldsymbol{\alpha}^{(t)}|+k$ with $z = |\boldsymbol{\alpha}|+k$.

(7) The inequality used in Zhou, et al. (2011) (p.269), i.e.,

$$(\lambda_j-\lambda_k)^2 \leqslant \frac{1}{2}(2\lambda_j - \lambda_j^{(t)} - \lambda_k^{(t)})^2 + \frac{1}{2}(2\lambda_k - \lambda_j^{(t)} - \lambda_k^{(t)})^2$$

is a special case of arithmetic geometric mean inequality when we rearrange the term

$$(\lambda_j-\lambda_k)^2 = \left[\frac{1}{2}(2\lambda_j - \lambda_j^{(t)} - \lambda_k^{(t)}) - \frac{1}{2}(2\lambda_k - \lambda_j^{(t)} - \lambda_k^{(t)})\right]^2.$$

The basic idea of an MM algorithms is that instead of maximizing the log-likelihood function, one must find a minorizing/surrogate function, which is maximized at each iteration. In this paper, we first proposed a new AD technique to construct separable minorizing functions in a class of MM algorithms, where in the A-technique, the notions of assemblies and complemental assemblies are introduced and in the D-technique, the log-likelihood function is decomposed into the sum of concave and/or convex functions under the guideline of the A-technique. Second, the applications of the proposed AD method to diverse applications are presented and new MM algorithms are developed, which were not previously reported in the literature. Third, the further extensions of the proposed AD technique were also considered.

When the conditions $a_j \geqslant 0,\ \theta_j \geqslant 0, j = 1, \ldots, q$, in (S7.17) are violated, i.e., if $a_j \in \mathbb{R}$ and $\theta_j \in \mathbb{R}$, for $j = 1, \ldots, q$, we could employ De Pierro's Algorithm (De Pierro (1995)) to calculate the MLE of $\boldsymbol{\theta}$. Take $\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = g_1(\boldsymbol{a}^\top\boldsymbol{\theta}) = c_0 + c_1 \log(\boldsymbol{a}^\top\boldsymbol{\theta}) - c_2(\boldsymbol{a}^\top\boldsymbol{\theta})^k$ for example, we construct weight $w_j = |a_j|/\sum_{j=1}^q |a_j|$ and rewrite $\boldsymbol{a}^\top\boldsymbol{\theta} = \sum_{j=1}^q w_j[w_j^{-1}a_j(\theta_j - \theta_j^{(t)}) +$

$\boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}]$, according to Jensen's inequality, we can obtain

$$\begin{aligned} g_1(\boldsymbol{a}^\top\boldsymbol{\theta}) &\geqslant c_0 + \sum_{j=1}^{q}\Big\{c_1 w_j \log\big[w_j^{-1}a_j(\theta_j - \theta_j^{(t)}) + \boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}\big] \\ &\quad -c_2 w_j\big[w_j^{-1}a_j(\theta_j - \theta_j^{(t)}) + \boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}\big]^k\Big\}, \\ &\hat{=} c_0 + \sum_{j=1}^{q} G_{1j}\big[w_j^{-1}a_j(\theta_j - \theta_j^{(t)}) + \boldsymbol{a}^\top\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}\big] \hat{=} Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}). \end{aligned}$$

References

Chi, E.C. and Lange, K. (2014). A look at the generalized heron problem through the lens of majorization–minimization. *The American Mathematical Monthly* **121**(2), 95–108.

De Pierro, A.R. (1995). A modified EM algorithm for penalized likelihood estimation in emission tomography. *IEEE Transactions on Medical Imaging* **14**, 132–137.

Lange, K. (2010). Numerical Analysis for Statisticians, second ed. *Statistics and Computing.* Springer, New York.

Lange, K. and Zhou, H. (2014). MM algorithms for geometric and signomial programming. *Mathematical Programming, Series A* **143**, 339–356.

Yen, T.J. (2011). A majorization–minimization approach to variable selection using spike and slab priors. *The Annals of Statistics* **39**(3), 1748–1775.

Zhou, H. and Zhang, Y.W. (2012). EM vs MM: A case study. *Computational Statistics and Data Analysis* **56**, 3909–3920.

Zhou, H., Alexander, D. and Lange, K. (2011). A quasi-Newton acceleration for highdimensional optimization algorithms. *Statistics and Computing* **21**(2), 261–273.