1 Le modèle

On a un modèle génératif M complexe qui se décompose en trois morceaux $M = (M_1, M_2, M_3)$ comme suit :

$$\begin{aligned}
M_1 : \quad & Y|X^{I_1}, X^{I_2}, Z \sim X^{I_1}A_{I_1} + X^{I_2}A_{I_2} + \varepsilon & (1)\\
M_2 : \quad & X^{I_2}|X^{I_1}, Z \sim X^{I_1}B_{I_1}^{I_2} + \varepsilon_{X^{I_2}} & (2)\\
M_3 : \quad & X^{I_1}|Z \sim GM(\alpha) & (3)
\end{aligned}$$

On pose $\theta = (\alpha, \Sigma, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2})$ Pour un Z donné on a donc

$$\begin{aligned}
P_{M,Z}(Y, X^{I_1}, X^{I_2}|\theta) &= P_{M,Z}(X^{I_1}|\theta)P_{M,Z}(X^{I_2}|X^{I_1};\theta)P_{M,Z}(Y|X^{I_2}, X^{I_1};\theta) & (4)\\
&= P_{M_3,Z}(X^{I_1}|\alpha)P_{M_1,M_2,Z}(X^{I_2}|X^{I_1};\Sigma,\Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2})P_{M_1,M_2,Z}(Y|X^{I_2}, X^{I_1};\Sigma,\Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2}) & (5)
\end{aligned}$$

$P_{M_3,Z}(X^{I_1}|\alpha)$ vit tout seul et est estimé par Mixmod. On sait également que

$$\begin{aligned}
P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma,\Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2}) &= P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma_{X^{I_2}}, B_{I_1}^{I_2}) \text{ et} & (6)\\
P_{M_1,Z}(Y|X^{I_2}, X^{I_1};\Sigma,\Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2}) &= P_{M_1,Z}(Y|X^{I_2}, X^{I_1};\Sigma, A_{I_1}, A_{I_2}) & (7)
\end{aligned}$$

De manière générale les gens s'arrêtent là, et obtiennent le modèle complet... qui ne tient pas compte de la structure. Mais nous, nous ne sommes pas juste sous M_2 pour les sous-régression ni juste sous M_1 pour Y. Donc la dépendance persiste (c'est ce qui permet que la contrainte aie une utilité, sinon elle serait inutile).

Du coup, il faut définir proprement $P_{M_1,M_2,Z}(X^{I_2}|X^{I_1};\Sigma,\Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2})$ et $P_{M_1,M_2,Z}(Y|X^{I_2}, X^{I_1};\Sigma,\Sigma_{X^{I_2}}, B_{I_1}^{I_2}, A_{I_1}, A_{I_2})$. Or il semble que ces vraisemblances ne soient pas explicites.

Vu que $P_{M_1,M_2,Z}(X^{I_2}|X^{I_1};\theta) \neq P_{M_2,Z}(X^{I_2}|X^{I_1};\theta)$ et $P_{M_1,M_2,Z}(Y|X^{I_2}, X^{I_1};\theta) \neq P_{M_1,Z}(Y|X^{I_2}, X^{I_1};\theta)$, on se limite donc à une décomposition partielle de la vraisemblance et on obtient :

$$P_{M,Z}(Y, X^{I_1}, X^{I_2}|\theta) = P_{M_3,Z}(X^{I_1}|\alpha)P_{M_1,M_2,Z}(Y, X^{I_2}|X^{I_1};\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}) \tag{8}$$

Le second membre suit une loi normale mulivariée :

$$(Y, X^{I_2}|X^{I_1}) \sim \mathcal{N}\left(\begin{pmatrix} X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}; \begin{pmatrix} \sigma_Y^2 + \sum_{i\in I_2}(\sigma_i^2 a_i^2) & \dots & a_j\sigma_j^2 & \dots \\ \vdots & \ddots & 0 & 0 \\ a_j\sigma_j^2 & 0 & \sigma_j^2 & 0 \\ \vdots & 0 & 0 & \ddots \end{pmatrix}\right) = \mathcal{N}\left(\begin{pmatrix} X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}; \begin{pmatrix} \sigma_Y^2 + \sum_{i\in I_2}(\sigma_i^2 a_i^2) & \dots a_j\sigma_j^2 \dots \\ \vdots & \\ a_j\sigma_j^2 & \Sigma_X^{I_2} \\ \vdots & \end{pmatrix}\right) \tag{9}$$

On note $\bar{\Sigma}$ la matrice de variance-covariance correspondante Avec par conséquent une vraisemblance de la forme :

$$P_{M_1,M_2,Z}(Y, X^{I_2}|X^{I_1}; \Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}) = \mathcal{L}_{M_1,M_2,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}); Y, X^{I_2}|X^{I_1}) \tag{10}$$

$$= \frac{1}{(2\pi)^{\frac{p_2+1}{2}}|\bar{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}^t \bar{\Sigma}^{-1} \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}\right) \tag{11}$$

avec $|\bar{\Sigma}|$ le déterminant de $\bar{\Sigma}$.

On passe à la log-vraisemblance :

$$L_{M_1,M_2,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}); Y, X^{I_2}|X^{I_1}) = -\frac{1}{2}\left[(p_2+1)\ln(2\pi) + \ln(|\bar{\Sigma}|) + \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}^t \bar{\Sigma}^{-1} \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}\right] \tag{12}$$

2 estimation

On veut maximiser la vraisemblance sous la contrainte $A_{I_1} + B_{I_1}^{I_2}A_{I_2} = \hat{\hat{A}}_{I_1}$ où $\hat{\hat{A}}_{I_1}$ a été préalablement calculé lors de l'étape explicative. Cette contrainte n'est pas linéaire mais devient linéaire si on se limite à A ou à B. On va donc procéder par optimisation alternée, en maximisant la vraisemblance pour un B fixé puis en recommençant avec le A obtenu, et ainsi de suite.La contrainte étant d'égalité, on utilisera le théorème de Lagrange. La partie de la vraisemblance relative à X^{I_1} ne dépend pas de la contrainte ni de l'autre membre et on la maximise donc en amont par Mixmod.

On commence par expliciter la matrice $\bar{\Sigma}^{-1}$

$$\bar{\Sigma} = \begin{pmatrix} E & C^t \\ C & \Sigma_{X^{I_2}} \end{pmatrix} \tag{13}$$

$$E = \sigma_Y^2 + \sum_{i \in I_2}(\sigma_i^2 A_i^2) = \sigma_Y^2 + A_{I_2}^t \Sigma_{X^{I_2}} A_{I_2} \text{ scalaire} \tag{14}$$

$$C = \begin{pmatrix} \vdots \\ a_i \sigma_i^2 \\ \vdots \end{pmatrix} = \Sigma_{X^{I_2}} A_{I_2} \text{ de taille } p_2 \times 1 \tag{15}$$

$$C^t = A_{I_2}^t \Sigma_{X^{I_2}}^t = A_{I_2}^t \Sigma_{X^{I_2}} \text{ de taille } 1 \times p_2 \tag{16}$$

On a donc à l'aide du complément de Schur :

$$\bar{\Sigma}^{-1} = \begin{pmatrix} [E - C^t\Sigma_{X^{I_2}}^{-1}C]^{-1} & . \\ -\Sigma_{X^{I_2}}^{-1}C(E - C^t\Sigma_{X^{I_2}}^{-1}C)^{-1} & [\Sigma_{X^{I_2}}^{-1} + \Sigma_{X^{I_2}}^{-1}C(E - C^t\Sigma_{X^{I_2}}^{-1}C)^{-1}C^t\Sigma_{X^{I_2}}^{-1}] \end{pmatrix} \text{ matrice symétrique} \tag{17}$$

Le complément de Schur est récurrent, on le note $S = E - C^t\Sigma_{X^{I_2}}^{-1}C$. Après calculs on obtient $S^{-1} = \frac{1}{\sigma_Y^2}$ On obtient alors :

$$-\Sigma_{X^{I_2}}^{-1}C(E - C^t\Sigma_{X^{I_2}}^{-1}C)^{-1} = -\Sigma_{X^{I_2}}^{-1}CS^{-1} = -\Sigma_{X^{I_2}}^{-1}\Sigma_{X^{I_2}}A_{I_2}S^{-1} = -\frac{1}{\sigma_Y^2}A_{I_2} \text{ de taille } p_2 \times 1 \tag{18}$$

et enfin :

$$[\Sigma_{X^{I_2}}^{-1} + \Sigma_{X^{I_2}}^{-1}C(E - C^t\Sigma_{X^{I_2}}^{-1}C)^{-1}C^t\Sigma_{X^{I_2}}^{-1}] = \Sigma_{X^{I_2}}^{-1} + \frac{1}{\sigma_Y^2}A_{I_2}A_{I_2}^t\Sigma_{X^{I_2}}\Sigma_{X^{I_2}}^{-1} = \Sigma_{X^{I_2}}^{-1} + \frac{1}{\sigma_Y^2}A_{I_2}A_{I_2}^t \text{ matrice symétrique pleine} \tag{19}$$

On a donc :

$$\bar{\Sigma}^{-1} = \begin{pmatrix} \frac{1}{\sigma_Y^2} & -\frac{1}{\sigma_Y^2}A_{I_2}^t \\ -\frac{1}{\sigma_Y^2}A_{I_2} & [\Sigma_{X^{I_2}}^{-1} + \frac{1}{\sigma_Y^2}A_{I_2}A_{I_2}^t] \end{pmatrix} \text{ matrice symétrique pleine} \tag{20}$$

On note $\Sigma_{X^{I_2}(-k)}$ la matrice $\Sigma_X^{I_2}$ dépourvue de la ligne$k$ et de la colonne k. On note $\bar{\Sigma}(-1,-k)$ la matrice $\bar{\Sigma}$ dépourvue de sa prémière colonne et de sa ligne k. On a alors

$$\begin{aligned}
|\bar{\Sigma}| &= (\sigma_Y^2 + A_{I_2}^t\Sigma_{X^{I_2}}A_{I_2})|\Sigma_X^{I_2}| + \sum_{k=1}^{p_2}(-1)^k A_{I_2(k)}\sigma_{I_2(k)}^2|\bar{\Sigma}(-1,-(k+1))| \text{ développement 1 colonne} && (21)\\
&= (\sigma_Y^2 + A_{I_2}^t\Sigma_{X^{I_2}}A_{I_2})|\Sigma_X^{I_2}| + \sum_{k=1}^{p_2}(-1)^k A_{I_2(k)}\sigma_{I_2(k)}^2[(-1)^{k+1}A_{I_2(k)}\sigma_{I_2(k)}^2\frac{1}{\sigma_{I_2(k)}^2}\prod_{j\in I_2}\sigma_j^2] \text{ developpements k colonne} && (22)\\
&= \sigma_Y^2\prod_{j\in I_2}\sigma_j^2 + \sum_{k=1}^{p_2}A_{I_2(k)}^2\sigma_{I_2(k)}^2\prod_{j\in I_2}\sigma_j^2 - \sum_{k=1}^{p_2}A_{I_2(k)}^2\sigma_{I_2(k)}^2\prod_{j\in I_2}\sigma_j^2 && (23)\\
&= \sigma_Y^2\prod_{j\in I_2}\sigma_j^2 && (24)
\end{aligned}$$

2.1 initialisation

On commence par choisir une valeur de B. On pourrait en prendre une arbitraire, mais on choisit de commencer par le B obtenu par estimation indépendante des sous-régressions (par maximisation de $P_{M_2,Z}(X^{I_2}|X^{I_1};\Sigma_{X^{I_2}},B_{I_1}^{I_2})$). Cela peut se faire par moindres carrés classiques appliqués de manière indépendante à chacune des sous-régressions.

2.2 étape A

pour cette étape, les paramètres $\Sigma_{X^{I_2}}$ et $B_{I_1}^{I_2}$) sont fixés (la connaissance de $B_{I_1}^{I_2}$) permet d'obtenir directement le $\Sigma_{X^{I_2}}$ associé. On se ramène à la maximisation de

$$\begin{aligned} \max_A f(A) &= \max_A L_{M_1,M_2,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B_{I_1}^{I_2}); Y, X^{I_2}|X^{I_1}) && (25)\\ s.c.\;\; A_{I_1} + B_{I_1}^{I_2} A_{I_2} - \hat{\hat{A}}_{I_1} &= 0 && (26) \end{aligned}$$

En notant $\psi(A) = A_{I_1} + B_{I_1}^{I_2} A_{I_2} - \hat{\hat{A}}_{I_1}$ la contrainte, le théorème de Lagrange permet de dire qu'il existe un unique λ pour lequel résoudre notre maximisation sous contrainte revient à résoudre :

$$g(A,\lambda) = \bigtriangledown f(A) + \lambda \bigtriangledown \psi(A) = 0 \tag{27}$$

pour $i \in I_1$, on remarque que $\bar{\Sigma}$ ne dépend pas de A_{I_1} :

$$\begin{aligned} \frac{\partial g(A,\lambda)}{\partial A_i} &= -\frac{1}{2}\left[\begin{pmatrix} X^i \\ 0 \end{pmatrix}^t \bar{\Sigma}^{-1} \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2} A_{I_2}) \\ (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t \end{pmatrix} + \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2} A_{I_2}) \\ (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t \end{pmatrix}^t \bar{\Sigma}^{-1} \begin{pmatrix} X^i \\ 0 \end{pmatrix}\right] + \lambda_i && (28)\\ &= -\frac{1}{2}\left[\begin{pmatrix} \frac{X^i}{\sigma_Y^2} \\ -\frac{X^i A_{I_2}}{\sigma_Y^2} \end{pmatrix}^t \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2} A_{I_2}) \\ (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t \end{pmatrix} + \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2} A_{I_2}) \\ (X^{I_2} - X^{I_1} B_{I_1}^{I_2})^t \end{pmatrix}^t \begin{pmatrix} \frac{X^i}{\sigma_Y^2} \\ -\frac{X^i A_{I_2}}{\sigma_Y^2} \end{pmatrix}\right] + \lambda_i && (29)\\ &= -\frac{1}{\sigma_Y^2}\left[X^i(Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2} A_{I_2}) - (X^{I_2} - X^{I_1} B_{I_1}^{I_2}) A_{I_2})\right] + \lambda_i && (30)\\ &= -\frac{1}{\sigma_Y^2}\left[X^i(Y - X^{I_1} A_{I_1} - X^{I_2} A_{I_2} - X^{I_1} B_{I_1}^{I_2} A_{I_2} + X^{I_1} B_{I_1}^{I_2} A_{I_2})\right] + \lambda_i && (31)\\ &= -\frac{1}{\sigma_Y^2}\left[X^i(Y - X^{I_1} A_{I_1} - X^{I_2} A_{I_2})\right] + \lambda_i && (32) \end{aligned}$$

pour $i \in I_2$:

$$\begin{aligned} \frac{\partial \bar{\Sigma}^{-1}}{\partial A_i} &= \frac{1}{\sigma_Y^2} \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & & & \\ -1 & & D & \\ 0 & & & \end{pmatrix} \text{ avec, pour i fixé} && (33)\\ D_{i,i} &= 2a_i && (34)\\ D_{i,j} &= a_j && (35)\\ D_{j,i} &= a_j && (36)\\ D &= 0 \text{ sinon} && (37) \end{aligned}$$

et on a :

$$\frac{\partial\bar{\Sigma}^{-1}}{\partial A_i}\begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix} = \frac{1}{\sigma_Y^2}\begin{pmatrix} -X^i + X^{I_1}B_{I_1}^i \\ (X^i - X^{I_1}B_{I_1}^i)A_{I_2} \end{pmatrix} + \frac{1}{\sigma_Y^2}\begin{pmatrix} 0 \\ -(Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2})) + (X^{I_2} - X^{I_1}B_{I_1}^{I_2})A_{I_2} \\ 0 \end{pmatrix} \quad (38)$$

$$= \frac{1}{\sigma_Y^2}(X^i - X^{I_1}B_{I_1}^i)\begin{pmatrix} -1 \\ A_{I_2} \end{pmatrix} + \frac{1}{\sigma_Y^2}\begin{pmatrix} 0 \\ -Y + X^{I_1}A_{I_1} + X^{I_2}A_{I_2} \\ 0 \end{pmatrix} \text{- > ligne i+1} \quad (39)$$

Donc la dérivée partielle de g aura un terme en

$$\begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix}^t \left(\frac{1}{\sigma_Y^2}(X^i - X^{I_1}B_{I_1}^i)\begin{pmatrix} -1 \\ A_{I_2} \end{pmatrix} + \frac{1}{\sigma_Y^2}\begin{pmatrix} 0 \\ -Y + X^{I_1}A_{I_1} + X^{I_2}A_{I_2} \\ 0 \end{pmatrix}\right) \quad (40)$$

$$= \frac{1}{\sigma_Y^2}(X^i - X^{I_1}B_{I_1}^i)[-(Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2})) + (X^{I_2} - X^{I_1}B_{I_1}^{I_2})A_{I_2}] + \frac{1}{\sigma_Y^2}(X^i - X^{I_1}B_{I_1}^i)(-Y + X^{I_1}A_{I_1} + X^{I_2}A_{I_2}) \quad (41)$$

$$= \frac{2}{\sigma_Y^2}(X^i - X^{I_1}B_{I_1}^i)[-Y + X^{I_1}A_{I_1} + X^{I_2}A_{I_2}] \quad (42)$$

$$\frac{\partial g(A,\lambda)}{\partial A_i} = -\frac{1}{2}\left[2\begin{pmatrix} A_iX^{I_1}B_{I_1}^i \\ 0 \end{pmatrix}^t \bar{\Sigma}^{-1}\begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix} + \frac{2}{\sigma_Y^2}(X^i - X^{I_1}B_{I_1}^i)[-Y + X^{I_1}A_{I_1} + X^{I_2}A_{I_2}]\right] + \sum_{j\in I_1}\lambda_j B_{j,i} \quad (43)$$

$$= -\frac{1}{\sigma_Y^2}\left[\begin{pmatrix} A_iX^{I_1}B_{I_1}^i \\ -A_iX^{I_1}B_{I_1}^iA_{I_2} \end{pmatrix}^t \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2}) \\ (X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t \end{pmatrix} + (X^i - X^{I_1}B_{I_1}^i)[-Y + X^{I_1}A_{I_1} + X^{I_2}A_{I_2}]\right] + \sum_{j\in I_1}\lambda_j B_{j,i} \quad (44)$$

$$= -\frac{1}{\sigma_Y^2}\left[A_iX^{I_1}B_{I_1}^i\left((Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2})) - A_{I_2}^t(X^{I_2} - X^{I_1}B_{I_1}^{I_2})^t\right) + (X^i - X^{I_1}B_{I_1}^i)[-Y + X^{I_1}A_{I_1} + X^{I_2}A_{I_2}]\right] + \sum_{j\in I_1}\lambda_j B_{j,i} \quad (45)$$

$$= -\frac{1}{\sigma_Y^2}\left[A_iX^{I_1}B_{I_1}^i\left((Y - X^{I_1}(A_{I_1} + B_{I_1}^{I_2}A_{I_2})) - (X^{I_2} - X^{I_1}B_{I_1}^{I_2})A_{I_2}\right) + (X^i - X^{I_1}B_{I_1}^i)[-Y + X^{I_1}A_{I_1} + X^{I_2}A_{I_2}]\right] + \sum_{j\in I_1}\lambda_j B_{j,i} \quad (46)$$

$$= -\frac{1}{\sigma_Y^2}\left[\left(Y - X^{I_1}A_{I_1} - X^{I_2}A_{I_2}\right)(A_iX^{I_1}B_{I_1}^i - X^i + X^{I_1}B_{I_1}^i)\right] + \sum_{j\in I_1}\lambda_j B_{j,i} \quad (47)$$

$$(48)$$

et enfin

$$\frac{\partial g(A,\lambda)}{\partial\lambda_i} = \frac{\partial\lambda\psi(A)}{\partial\lambda_i} = \lambda_i(B_iA - \hat{\hat{A}}_i) = \lambda_i(\sum_{j\in I_2}B_{i,j}A_j - \hat{\hat{A}}_i) \quad (49)$$

On a donc un système à résoudre de la forme

$$\begin{aligned}
\lambda_i(\sum_{j\in I_2} B_{i,j}A_j - \hat{\hat{A}}_i) &= 0 \ \forall i \in I_1 &(50)\\
-\frac{1}{\sigma_Y^2}\left[X^i(Y - X^{I_1}A_{I_1} - X^{I_2}A_{I_2})\right] + \lambda_i &= 0 \ \forall i \in I_1 &(51)\\
-\frac{1}{\sigma_Y^2}\left[\left(Y - X^{I_1}A_{I_1} - X^{I_2}A_{I_2}\right)(A_i X^{I_1}B^i_{I_1} - X^i + X^{I_1}B^i_{I_1})\right] + \sum_{j\in I_1}\lambda_j B_{j,i} &= 0 \ \forall i \in I_2 &(52)
\end{aligned}$$

2.3 étape B

Pour cette étape, les paramètres Σ, A_{I_1} et A_{I_2} sont fixés (ici aussi, la connaissance de A permet d'obtenir le Σ associé). Comme pour l'étape A, on va résoudre :

$$\bigtriangledown f(B) + \lambda \bigtriangledown \psi(B) = 0 \tag{53}$$

où $f(B) = L_{M_1,M_2,Z}(\Sigma, A_{I_1}, A_{I_2}, \Sigma_{X^{I_2}}, B^{I_2}_{I_1}); Y, X^{I_2}|X^{I_1})$

$$g(B,\lambda) = \bigtriangledown f(B) + \lambda \bigtriangledown \psi(B) = 0 \tag{54}$$

pour $i \in I_1, j \in I_2$ et $Z_{i,j} \neq 0$:

$$\frac{\partial g(B,\lambda)}{\partial B_{i,j}} = -\frac{1}{2}\left[\begin{pmatrix} X^iA_j \\ 0 \\ X^i \\ 0 \end{pmatrix}^t \bar{\Sigma}^{-1}\begin{pmatrix} Y - X^{I_1}(A_{I_1} + B^{I_2}_{I_1}A_{I_2}) \\ (X^{I_2} - X^{I_1}B^{I_2}_{I_1})^t \end{pmatrix} + \begin{pmatrix} Y - X^{I_1}(A_{I_1} + B^{I_2}_{I_1}A_{I_2}) \\ (X^{I_2} - X^{I_1}B^{I_2}_{I_1})^t \end{pmatrix}^t \bar{\Sigma}^{-1}\begin{pmatrix} X^iA_j \\ 0 \\ X^i \\ 0 \end{pmatrix}\right] + \lambda_i a_j \tag{55}$$

$$\frac{\partial g(B,\lambda)}{\partial \lambda_i} = \frac{\partial \lambda\psi(B)}{\partial \lambda_i} = \lambda_i(B_iA - \hat{\hat{A}}_i) = \lambda_i(\sum_j B_{i,j}A_j - \hat{\hat{A}}_i) \tag{56}$$