1 Model 1

Suppose we have cell type data C for n_1 of the individuals and no cell type data for the other n_2 individuals. The models is then

$$Y_1 = BX_1 + LC + \Gamma\alpha X_1 + \Gamma W_1 + E_1$$

$$Y_2 = BX_2 + LD\Xi + L\Omega X_2 + \Gamma\alpha X_2 + \Gamma W_2 + E_2$$

where $DD^T = \Lambda$, $W_i \sim MN_{r\times n_i}(0, I_r, I_{n_i})$, $\Xi \sim MN_{K\times n_2}(0, I_K, I_{n_2})$, $E_i \sim MN_{p\times n_i}(0, \Sigma, I_{n_i})$. Let $X_i^T = Q_i R_i$, $Q_i = \begin{bmatrix} Q_i^{(1)} & Q_i^{(2)} \end{bmatrix}$, $Z_i = Y_i Q_i$. Then

$$Z_1^{(1)} = (B + \Gamma\alpha) R_1 + L\underbrace{CQ_1^{(1)}}_{F_1} + \Gamma\tilde{W}_1^{(1)} + \tilde{E}_1^{(1)} \in \mathbb{R}^{p\times d} \quad (1)$$

$$Z_1^{(2)} = L\underbrace{CQ_1^{(2)}}_{F_2} + \Gamma\tilde{W}_1^{(2)} + \tilde{E}_1^{(2)} \in \mathbb{R}^{p\times(n_1-d)} \quad (2)$$

$$Z_2^{(1)} = (B + \Gamma\alpha + L\Omega) R_2 + LD\tilde{\Xi}^{(1)} + \Gamma\tilde{W}_2^{(1)} + \tilde{E}_2^{(1)} \in \mathbb{R}^{p\times d} \quad (3)$$

$$Z_2^{(2)} = LD\tilde{\Xi}^{(2)} + \Gamma\tilde{W}_2^{(2)} + \tilde{E}_2^{(2)} \in \mathbb{R}^{p\times(n_2-d)} \quad (4).$$

I will use (2) and (4) to estimate $L, \Gamma, \Lambda, \Sigma$.

1.1 Preliminary Identities

Let $\Psi = \Sigma + \Gamma\Gamma^T$ and $J = L\Lambda L^T + \Psi$. For notation purposes, I will let $Z_1 = Z_1^{(2)}$, $Z_2 = Z_2^{(2)}$, $S_i = Z_i Z_i^T$, $F = F_2$ and $m_i = n_i - d$.

$$\Psi^{-1} = \left(\Sigma + \Gamma\Gamma^T\right)^{-1} = \Sigma^{-1} - \Sigma^{-1}\Gamma\left(I_r + \Gamma^T\Sigma^{-1}\Gamma\right)^{-1}\Gamma^T\Sigma^{-1}$$

$$J^{-1} = \left(\Psi + L\Lambda L^T\right)^{-1} = \Psi^{-1} - \Psi^{-1}L\left(\Lambda^{-1} + L^T\Psi^{-1}L\right)^{-1}L^T\Psi^{-1}$$

For the subsequent analysis, I need to be able to compute the following matrices:

1.
$$\text{LtSinvZ} = L^T\Sigma^{-1}Z_2$$

2.
$$\text{LtSinvG} = L^T\Sigma^{-1}\Gamma$$

3.
$$\text{GtSinvG} = \Gamma^T\Sigma^{-1}\Gamma$$

4.
$$\text{GtSinvZ} = \Gamma^T\Sigma^{-1}Z_2$$

5.
$$\text{LtSinvL} = L^T\Sigma^{-1}L$$

6.
$$\text{AtSinvG} = A^T\Sigma^{-1}\Gamma$$

7.
$$\text{mid.r} = I_r + \text{GtSinvG}$$

8.
$$\text{mid.Lambda} = \Lambda^{-1} + H_2$$

Using the above Woodbury identities, we need to compute the following matrices:

1.
$$H_1 = L^T \Psi^{-1} Z_2 = L^T \Sigma^{-1} Z_2 - L^T \Sigma^{-1} \Gamma \left(I_r + \Gamma^T \Sigma^{-1} \Gamma\right)^{-1} \Gamma^T \Sigma^{-1} Z_2$$

2.
$$H_2 = L^T \Psi^{-1} L = L^T \Sigma^{-1} L - L^T \Sigma^{-1} \Gamma \left(I_r + \Gamma^T \Sigma^{-1} \Gamma\right)^{-1} \Gamma^T \Sigma^{-1} L$$

3.
$$H_3 = A^T \Psi^{-1} \Gamma = A^T \Sigma^{-1} \Gamma - A^T \Sigma^{-1} \Gamma \left(I_r + \Gamma^T \Sigma^{-1} \Gamma\right)^{-1} \Gamma^T \Sigma^{-1} \Gamma$$

4.
$$H_4 = Z_2^T \Psi^{-1} \Gamma = Z_2^T \Sigma^{-1} \Gamma - Z_2^T \Sigma^{-1} \Gamma \left(I_r + \Gamma^T \Sigma^{-1} \Gamma\right)^{-1} \Gamma^T \Sigma^{-1} \Gamma$$

5.
$$H_5 = L^T \Psi^{-1} \Gamma = L^T \Sigma^{-1} \Gamma - L^T \Sigma^{-1} \Gamma \left(I_r + \Gamma^T \Sigma^{-1} \Gamma\right)^{-1} \Gamma^T \Sigma^{-1} \Gamma$$

6.
$$H_6 = \Gamma^T \Psi^{-1} \Gamma = \Gamma^T \Sigma^{-1} \Gamma - \Gamma^T \Sigma^{-1} \Gamma \left(I_r + \Gamma^T \Sigma^{-1} \Gamma\right)^{-1} \Gamma^T \Sigma^{-1} \Gamma$$

7.
$$H_7 = L^T \Psi^{-1} Z_1 = L^T \Sigma^{-1} Z_1 - L^T \Sigma^{-1} \Gamma \left(I_r + \Gamma^T \Sigma^{-1} \Gamma\right)^{-1} \Gamma^T \Sigma^{-1} Z_1$$

Using the above identities, we need to compute the following matrices:

1.
$$G_1 = L^T J^{-1} Z_2 = L^T \Psi^{-1} Z_2 - L^T \Psi^{-1} L \left(\Lambda^{-1} + L^T \Psi^{-1} L\right)^{-1} L^T \Psi^{-1} Z_2 = H_1 - H_2 \left(\Lambda^{-1} + H_2\right)^{-1} H_1$$
$$= H_1 - H_2 D \left(I_K + D^T H_2 D\right)^{-1} D^T H_1$$

which is used in $L, \Lambda, \Gamma, \Sigma$.

2.
$$G_2 = L^T J^{-1} L = L^T \Psi^{-1} L - L^T \Psi^{-1} L \left(\Lambda^{-1} + L^T \Psi^{-1} L\right)^{-1} L^T \Psi^{-1} L = H_2 - H_2 \left(\Lambda^{-1} + H_2\right)^{-1} H_2$$
$$= H_2 - H_2 D \left(I_K + D^T H_2 D\right)^{-1} D^T H_2$$

which is used in L, Λ, Σ.

3.
$$G_3 = A^T \Psi^{-1} \Gamma = H_3$$

which is used in Γ, Σ.

4.
$$G_4 = Z_2^T J^{-1} \Gamma = Z_2^T \Psi^{-1} \Gamma - Z_2^T \Psi^{-1} L \left(\Lambda^{-1} + L^T \Psi^{-1} L\right)^{-1} L^T \Psi^{-1} \Gamma = H_4 - H_1^T \left(\Lambda^{-1} + H_2\right)^{-1} H_5$$
$$= H_4 - H_1^T D \left(I_K + D^T H_2 D\right)^{-1} D^T H_5$$

which is used in Γ, Σ.

5.
$$G_5 = L^T J^{-1} \Gamma = L^T \Psi^{-1} \Gamma - L^T \Psi^{-1} L \left(\Lambda^{-1} + L^T \Psi^{-1} L\right)^{-1} L^T \Psi^{-1} \Gamma = H_5 - H_2 \left(\Lambda^{-1} + H_2\right)^{-1} H_5$$
$$= H_5 - H_2 D \left(I_K + D^T H_2 D\right)^{-1} D^T H_5$$

which is used in Γ, Σ.

6.
$$G_6 = \Gamma^T \Psi^{-1} \Gamma = H_6$$

which is used in Γ, Σ.

7.
$$G_7 = \Gamma^T J^{-1} \Gamma = \Gamma^T \Psi^{-1} \Gamma - \Gamma^T \Psi^{-1} L \left(\Lambda^{-1} + L^T \Psi^{-1} L\right)^{-1} L^T \Psi^{-1} \Gamma = H_6 - H_5^T \left(\Lambda^{-1} + H_2\right)^{-1} H_5$$
$$= H_6 - H_5^T D \left(I_K + D^T H_2 D\right)^{-1} D^T H_5$$

which is used in Γ, Σ.

1.2   Update for L

Assume Γ, Λ, Σ are known. We then have

$$\tilde{\Xi}^{(2)} \mid Z_2 \sim MN_{K \times m_2}\left(D^T L^T J^{-1} Z_2, I_K - D^T L^T J^{-1} LD, I_{m_2}\right)$$

$$g_1(L) = \log f(Z_1 \mid L) = \frac{m_1}{2}\log\left|\Psi^{-1}\right| - \frac{1}{2}\mathrm{Tr}\left(\Psi^{-1}S_1\right) + \mathrm{Tr}\left(L^T\Psi^{-1}Y_1 F^T\right) - \frac{1}{2}\mathrm{Tr}\left(L^T\Psi^{-1}LFF^T\right)$$

$$g_2(L) = E_{\tilde{\Xi}^{(2)}\mid Z_2, L_t}\log f\left(Z_2 \mid L, \tilde{\Xi}^{(2)}\right) = \frac{m_2}{2}\left|\Psi^{-1}\right| - \frac{1}{2}\mathrm{Tr}\left(\Psi^{-1}S_2\right) + \mathrm{Tr}\left(L^T\Psi^{-1}S_2 J_t^{-1}L_t\Lambda\right)$$

$$- \frac{1}{2}\mathrm{Tr}\left[L^T\Psi^{-1}L\left(\Lambda L_t^T J_t^{-1}S_2 J_t^{-1}L_t\Lambda + m_2\left\{\Lambda - \Lambda L_t^T J_t^{-1}L_t\Lambda\right\}\right)\right]$$

$$\Rightarrow 0 = \frac{dg_1}{dL} + \frac{dg_2}{dL} = \Psi^{-1}Z_1 F^T - \Psi^{-1}LFF^T + \Psi^{-1}S_2 J_t^{-1}L_t\Lambda - \Psi^{-1}L\left(\Lambda L_t^T J_t^{-1}S_2 J_t^{-1}L_t\Lambda + m_2\left\{\Lambda - \Lambda L_t^T J_t^{-1}L_t\Lambda\right\}\right)$$

The update for L is then:

$$L = MU^{-1}$$

$$M = Z_1 F^T + S_2 J_t^{-1}L_t\Lambda = Z_1 F^T + Z_2 G_1^T\Lambda$$

$$U = FF^T + \Lambda L_t^T J_t^{-1}S_2 J_t^{-1}L_t\Lambda + m_2\left\{\Lambda - \Lambda L_t^T J_t^{-1}L_t\Lambda\right\} = FF^T + \Lambda G_1 G_1^T\Lambda + m_2\left\{\Lambda - \Lambda G_2\Lambda\right\}$$

I then need to update H_1, H_2, H_3, H_5 and then $G_1, G_2, G_3, G_4, G_5, G_7$ and then A.

1.3   Update for Λ

Assume that L, Γ, Σ are known. Let $V = D\tilde{\Xi}^{(2)} \sim MN_{K \times m_2}\left(0, \Lambda, I_{m_2}\right)$. Then

$$E_{V\mid Z_2, \Lambda_t}\log f(V \mid \Lambda) = \frac{m_2}{2}\log\left|\Lambda^{-1}\right| - \frac{1}{2}\mathrm{Tr}\left[\Lambda^{-1}\left(\Lambda_t L^T J_t^{-1}S_2 J_t^{-1}L\Lambda_t + m_2\left\{\Lambda_t - \Lambda_t L^T J_t^{-1}L\Lambda_t\right\}\right)\right]$$

The update for Λ is then

$$\Lambda = \frac{1}{m_2}\Lambda_t L^T J_t^{-1}S_2 J_t^{-1}L\Lambda_t + \Lambda_t - \Lambda_t L^T J_t^{-1}L\Lambda_t = \frac{1}{m_2}\Lambda_t G_1 G_1^T\Lambda_t + \Lambda_t - \Lambda_t G_2\Lambda_t$$

I then need to update G_1, G_2, G_4, G_5, G_7.

It turns out this does not work and often times gives updates that are not positive definite. The better option is to update D directly by using the second set of data:

$$\log f(Z_2 \mid V, D) = -\frac{m_2}{2}\log|\psi| - \frac{1}{2}\mathrm{Tr}\left[(Z_2 - LD\Xi_2)^T\Psi^{-1}(Z_2 - LD\Xi_2)\right] =$$

$$-\frac{m_2}{2}\log|\psi| - \frac{1}{2}\mathrm{Tr}\left[\Psi^{-1}Z_2 Z_2^T\right] + \mathrm{Tr}\left[\Xi_2^T D^T L^T\Psi^{-1}Z_2\right] - \frac{1}{2}\mathrm{Tr}\left[\Xi_2^T D^T L^T\Psi^{-1}LD\Xi_2\right]$$

Using this, the update is then

$$D = \left(L^T\Psi^{-1}L\right)^{-1}L^T\Psi^{-1}Z_2 E\left(\Xi_2^T \mid Z_2\right)E\left(\Xi_2\Xi_2^T \mid Z_2\right)^{-1} = MU^{\dagger}$$

$$M = H_2^{-1}H_1 G_1^T D_t$$

$$U = D_t^T G_1 G_1^T D_t + m_2\left\{I_K - D_t^T G_2 D_t\right\}$$

where is U is invertible, $U^{\dagger} = U^{-1}$. Note that $U^{\dagger} = VD_*^{\dagger}V^T$ where $U = VD_*V^T$ is the SVD of U.

This update for D also does not work. It may be best to optimize Λ directly from the likelihood function, given L, Γ, Σ. The likelihood function is

$$f(\Lambda) = -ll(\Lambda) = \frac{m_2}{2}\log\left|\Psi + L\Lambda L^T\right| + \frac{1}{2}\mathrm{Tr}\left[S_2\left(\Psi + L\Lambda L^T\right)^{-1}\right]$$

and the minimization we want to perform is

$$\min_{\Lambda \geq \epsilon} f\left(\Lambda\right) = \min_{\tilde{\Lambda} \geq 0} f\left(\tilde{\Lambda} + \epsilon\right) = \min_{\tilde{\Lambda} \geq 0} \tilde{f}\left(\tilde{\Lambda}\right), \text{ where } \tilde{\Lambda} = \Lambda - \epsilon$$

The Lagrangian for this problem is

$$\mathcal{L}\left(\tilde{\Lambda}, M\right) = \tilde{f}\left(\tilde{\Lambda}\right) - \text{Tr}\left(\tilde{\Lambda} M\right)$$

and

$$\nabla_{\tilde{\Lambda}} \mathcal{L}\left(\tilde{\Lambda}, M\right) = L^T J^{-1} L - \frac{1}{m_2} L^T J^{-1} S_2 J^{-1} L - M = G_2 - \frac{1}{m_2} G_1 G_1^T - M = g - M.$$

Where J is evaluated at Λ, NOT $\tilde{\Lambda}$. We know that we have reached a stable point when $G_2 - \frac{1}{m_2} G_1 G_1^T = M_* \geq 0$. Let $\tilde{\Lambda}_* = Q \tilde{D}_* Q^T$. If $M_* = 0$, there is nothing to compute. However, if $M_* \neq 0$, we must have

$$\text{Tr}\left(\tilde{D}_* \tilde{M}_*\right) = 0, \text{ where } \tilde{M}_* = Q^T M_* Q$$

meaning that if $\tilde{d}_{*_k} = 0$, then the kth row/column of $\tilde{M}_*$ must also me 0.

The Hessian is

$$H \underbrace{=}_{\text{column stacking}} \nabla^2_{\text{vec}(\tilde{\Lambda}) \text{vec}(\tilde{\Lambda})} \mathcal{L}\left(\tilde{\Lambda}, M\right) = \left(\frac{1}{m_2} G_1 G_1^T\right) \otimes \left(\frac{1}{m_2} G_1 G_1^T\right) - \left[G_2 - \frac{1}{m_2} G_1 G_1^T\right] \otimes \left[G_2 - \frac{1}{m_2} G_1 G_1^T\right]$$

where

$$EH = N = G_2 \otimes G_2.$$

The quadratic subproblem can then be stated as

$$\min_{x_k \geq 0} \frac{1}{2} \text{vec}\left(x_k - \tilde{\Lambda}_k\right)^T N \text{vec}\left(x_k - \tilde{\Lambda}_k\right) + \text{vec}\left(x_k - \tilde{\Lambda}_k\right)^T \text{vec}\left(g\right)$$

which has unconstrained update

$$\tilde{\Lambda}_{k+1} = x_k = \tilde{\Lambda}_k - N^{-1} \text{vec}\left(g\right) = \tilde{\Lambda}_k - G_2^{-1} \left(G_2 - \frac{1}{m_2} G_1 G_1^T\right) G_2^{-1} = \tilde{\Lambda}_k + \frac{1}{m_2} G_2^{-1} G_1 G_1^T G_2^{-1} - G_2^{-1}$$

It can be shown that the solution to the constrained problem with the approximate Hessian will ALWAYS decrease the function f, and is therefore a valid approximation. Unfortunately, Rmosek does not support such an optimization. Instead, I will use BFGS:

Let $\Lambda = RR^T$, $h_1\left(R\right) = \log\left|\Psi + LRR^T L^T\right|$ and $h_2\left(R\right) = \text{Tr}\left[S_2 \left(\Psi + LRR^T L^T\right)^{-1}\right]$. Note that $g = -ll = \frac{m_2}{2} h_1 + \frac{1}{2} h_2$. We also have

$$\frac{dh_1}{dR} = 2L^T J^{-1} LR = 2G_2 R$$

$$\frac{dh_2}{dR} = -2G_1 G_1^T R$$

and

$$\frac{dg}{dR} = \frac{m_2}{2}\left(G_2 R + R^T G_2\right) - \frac{1}{2}\left(G_1 G_1^T R + R^T G_1 G_1^T\right).$$

This can then be used as input into the function "optim".

1.4 Update for Γ

Assume that L, Λ, Σ are known. Define $A = Z_1 - LF$. Then

$$g_1\left(\Gamma\right) = E_{\tilde{W}_1^{(2)}|A,\Gamma_t} \log f\left(A \mid \tilde{W}_1^{(2)}\right) = \frac{m_1}{2} \log\left|\Sigma^{-1}\right| - \frac{1}{2} \text{Tr}\left(\Sigma^{-1} AA^T\right) + \text{Tr}\left(\Gamma^T \Sigma^{-1} AA^T \Psi_t^{-1} \Gamma_t\right) -$$

$$\frac{1}{2} \text{Tr}\left[\Gamma^T \Sigma^{-1} \Gamma \left(\Gamma_t^T \Psi_t^{-1} AA^T \Psi_t^{-1} \Gamma_t + m_1 \left\{I_r - \Gamma_t^T \Psi_t^{-1} \Gamma_t\right\}\right)\right]$$

For the next part, first note that

$$\begin{pmatrix} V \\ \tilde{W}_2^{(2)} \end{pmatrix} \mid Z_2 \sim MN_{(K+r)\times m_2}\left(\begin{pmatrix} \Lambda L^T \\ \Gamma^T \end{pmatrix} J^{-1} Z_2, \begin{pmatrix} \Lambda & 0 \\ 0 & I_r \end{pmatrix} - \begin{pmatrix} \Lambda L^T \\ \Gamma^T \end{pmatrix} J^{-1} \begin{pmatrix} L\Lambda & \Gamma \end{pmatrix}, I_{m_2}\right)$$

and that

$$\log f\left(Z_2 \mid V, \tilde{W}_2^{(2)}\right) = \frac{m_2}{2}\log|\Sigma^{-1}| - \frac{1}{2}\operatorname{Tr}\left(\Sigma^{-1}S_2\right) + \operatorname{Tr}\left(V^T L^T \Sigma^{-1} Z_2\right) + \operatorname{Tr}\left(\tilde{W}_2^{(2)}\Gamma^T \Sigma^{-1} Z_2\right) - \operatorname{Tr}\left(\tilde{W}_2^{(2)}\Gamma^T \Sigma^{-1} LV\right) - \frac{1}{2}\operatorname{Tr}\left(V^T L^T \Sigma^{-1} LV\right) -$$

$$\frac{1}{2}\operatorname{Tr}\left(\left(\tilde{W}_2^{(2)}\right)^T \Gamma^T \Sigma^{-1} \Gamma \tilde{W}_2^{(2)}\right).$$

Therefore,

$$g_2(\Gamma) = E_{\begin{pmatrix} V \\ \tilde{W}_2^{(2)} \end{pmatrix}|Z_2} \log f\left(Z_2 \mid V, \tilde{W}_2^{(2)}\right) = \frac{m_2}{2}\log|\Sigma^{-1}| - \frac{1}{2}\operatorname{Tr}\left(\Sigma^{-1}S_2\right) + \operatorname{Tr}\left(L^T \Sigma^{-1} S_2 J_t^{-1} L\Lambda\right) + \operatorname{Tr}\left(\Gamma^T \Sigma^{-1} S_2 J_t^{-1} \Gamma_t^T\right) -$$

$$\operatorname{Tr}\left(\Gamma^T \Sigma^{-1} L\left\{\Lambda L^T J_t^{-1} S_2 J_t^{-1} \Gamma_t - m_2 \Lambda L^T J_t^{-1} \Gamma_t\right\}\right) - \frac{1}{2}\operatorname{Tr}\left[L^T \Sigma^{-1} L\left(\Lambda L^T J_t^{-1} S_2 J_t^{-1} L\Lambda + m_2\left\{\Lambda - \Lambda L^T J_t^{-1} L\Lambda\right\}\right)\right] -$$

$$\frac{1}{2}\operatorname{Tr}\left[\Gamma^T \Sigma^{-1} \Gamma\left(\Gamma_t^T J_t^{-1} S_2 J_t^{-1} \Gamma_t + m_2\left\{I_r - \Gamma_t^T J_t^{-1} \Gamma_t\right\}\right)\right]$$

Setting $\frac{dg_1}{d\Gamma} + \frac{dg_2}{d\Gamma} = 0$, we find that the update for Γ is

$$\Gamma = MU^{-1}$$

$$M = AA^T \Psi_t^{-1} \Gamma_t + S_2 J_t^{-1} \Gamma_t - L\left\{\Lambda L^T J_t^{-1} S_2 J_t^{-1} \Gamma_t - m_2 \Lambda L^T J_t^{-1} \Gamma_t\right\} = AG_3 + Z_2 G_4 - L\{\Lambda G_1 G_4 - m_2 \Lambda G_5\}$$

$$U = \left(\Gamma_t^T \Psi_t^{-1} AA^T \Psi_t^{-1} \Gamma_t + m_1\left\{I_r - \Gamma_t^T \Psi_t^{-1} \Gamma_t\right\}\right) + \left(\Gamma_t^T J_t^{-1} S_2 J_t^{-1} \Gamma_t + m_2\left\{I_r - \Gamma_t^T J_t^{-1} \Gamma_t\right\}\right) = G_3^T G_3 + m_1\{I_r - G_6\} + G_4^T G_4 + m_2\{I_r - G_7\}.$$

I then need to update $H_1 - H_6$ and then $G_1 - G_7$.

1.5 Update for Σ

Assume L, Γ, Λ are known. Using g_1, g_2 from above and setting the derivative to 0, the update for $\Sigma = \operatorname{diag}(v)$ is

$$(m_1 + m_2)v = \operatorname{diag}\left(AA^T\right) + \operatorname{diag}(S_2) - 2\operatorname{diag}\left(AA^T \Psi^{-1}\Gamma\Gamma^T\right) - 2\operatorname{diag}\left(S_2 J^{-1}\Gamma\Gamma^T\right) + \operatorname{diag}\left[\Gamma\left(\Gamma^T\Psi^{-1}AA^T\Psi^{-1}\Gamma + m_1\left\{I_r - \Gamma^T\Psi^{-1}\Gamma\right\}\right)\Gamma^T\right]$$

$$+2\operatorname{diag}\left[L\left(\Lambda L^T J^{-1}S_2 J^{-1}\Gamma - m_2\Lambda L^T J^{-1}\Gamma\right)\Gamma^T\right] + \operatorname{diag}\left[\Gamma\left(\Gamma^T J^{-1}S_2 J^{-1}\Gamma + m_2\left\{I_r - \Gamma^T J^{-1}\Gamma\right\}\right)\Gamma^T\right]$$

$$+ \operatorname{diag}\left[L\left(\Lambda L^T J^{-1}S_2 J^{-1}L\Lambda + m_2\left\{\Lambda - \Lambda L^T J^{-1}L\Lambda\right\}\right)L^T\right] - 2\operatorname{diag}\left(S_2 J^{-1}L\Lambda L^T\right)$$

$$= \operatorname{diag}\left(AA^T\right) + \operatorname{diag}(S_2) + \operatorname{diag}\left[\left(\Gamma\left(G_3^T G_3 + m_1\{I_r - G_6\} + G_4^T G_4 + m_2\{I_r - G_7\}\right) + 2L\left(\Lambda G_1 G_4 - m_2\Lambda G_5\right) - 2AG_3 - 2Z_2 G_4\right)\Gamma^T\right]$$

$$+ \operatorname{diag}\left[\left(L\left(\Lambda G_1 G_1^T \Lambda + m_2\{\Lambda - \Lambda G_2\Lambda\}\right) - 2Z_2 G_1^T \Lambda\right)L^T\right].$$

I then need to update $H_1 - H_6$ and $G_1 - G_7$.

1.6 Compute Likelihood

The log-likelihood is

$$l(L, \Gamma, \Lambda, \Sigma \mid Z_1, Z_2) = -\frac{m_1}{2}\log|\Psi| - \frac{m_2}{2}\log|J| - \frac{1}{2}\operatorname{Tr}\left(A^T\Psi^{-1}A\right) - \frac{1}{2}\operatorname{Tr}\left(Z_2^T J Z_2\right) =$$

$$-\left(\frac{m_1 + m_2}{2}\right)\left(\log|I_r + \Gamma^T\Sigma^{-1}\Gamma| + \log|\Sigma|\right) - \frac{m_2}{2}\left[\log|\Lambda| + \log|\Lambda^{-1} + H_2|\right]$$

$$-\frac{1}{2}\operatorname{Tr}\left(A^T\Sigma^{-1}A\right) + \frac{1}{2}\operatorname{Tr}\left[A^T\Sigma^{-1}\Gamma\left(I_r + \Gamma^T\Sigma^{-1}\Gamma\right)^{-1}\Gamma^T\Sigma^{-1}A\right]$$

$$-\frac{1}{2}\operatorname{Tr}\left(Z_2^T\Sigma^{-1}Z_2\right) + \frac{1}{2}\operatorname{Tr}\left[Z_2^T\Sigma^{-1}\Gamma\left(I_r + \Gamma^T\Sigma^{-1}\Gamma\right)^{-1}\Gamma^T\Sigma^{-1}Z_2\right] + \frac{1}{2}\operatorname{Tr}\left[H_1^T\left(\Lambda^{-1} + H_2\right)^{-1}H_1\right]$$

5

1.7 BLUPs for V, $\tilde{W}_1^{(2)}$ and $\tilde{W}_2^{(2)}$

Using the above conditional distributions, see that

$$\hat{V} = E(V \mid Z_2) = \Lambda L^T J^{-1} Z_2 = \Lambda G_1$$

An estimate for C_2 is then

$$\hat{C}_2 = \hat{V}\left(Q_2^{(2)}\right)^T + \Omega X_2$$

$$\hat{\tilde{W}}_1^{(2)} = E\left(\tilde{W}_1^{(2)} \mid Z_1\right) = \Gamma^T \Psi^{-1}(Z_1 - LF) = G_3^T$$

$$\hat{\tilde{W}}_2^{(2)} = E\left(\tilde{W}_2^{(2)} \mid Z_2\right) = \Gamma^T J^{-1} Z_2 = G_4^T$$

These BLUPs DO NOT work! Another possibility is to get the GLS estimates for W and Ξ. Let $P_F^\perp$ be the orthogonal projection onto the subspace perpendicular to the rows of F. An estimate for $\tilde{W}_1^{(2)}$ is

$$\hat{\tilde{W}}_1^{(2)} = \left(\Gamma^T \Sigma^{-1} \Gamma\right)^{-1} \Gamma^T \Sigma^{-1} Z_1 P_F^\perp$$

Note that this is orthogonal to F. This does not matter, since we are only interested in estimating the subspace spanned by F and $\hat{\tilde{W}}_1^{(2)}$. By the same reasoning, estimates for V and $\tilde{W}_2^{(2)}$ are

$$\begin{bmatrix} \hat{V} \\ \hat{\tilde{W}}_2^{(2)} \end{bmatrix} = \begin{bmatrix} L^T \Sigma^{-1} L & L^T \Sigma^{-1} \Gamma \\ \Gamma^T \Sigma^{-1} L & \Gamma^T \Sigma^{-1} \Gamma \end{bmatrix}^{-1} \begin{bmatrix} L^T \Sigma^{-1} Z_2 \\ \Gamma^T \Sigma^{-1} Z_2 \end{bmatrix}$$

1.8 Starting Points for L, Λ and Γ

Below are starting points for L and Λ:

$$L_0 = Z_1 F^T \left(F F^T\right)^{-1}$$

$$\Lambda_0 = \frac{1}{m_1} F F^T$$

For Γ, let U_1 be such that $U_1 F^T = 0$ and $\tilde{Z}_1 = Z_1 U_1$. One possible starting point for Γ is

$$\text{1)} \quad \Gamma_0 = \frac{1}{\sqrt{m_1 - K}} U_{\tilde{Z}_1} D_{\tilde{Z}_1}$$

where $\tilde{Z}_1 = U_{\tilde{Z}_1} D_{\tilde{Z}_1} V_{\tilde{Z}_1}^T$ is the SVD of $\tilde{Z}_1$. The problem with this is that it only uses the data from the first set of individuals. Ideally we would use both to get a better estimate.

Another way to get a starting point would be to estimate F_2, the cell composition for the second set of individuals. Let $\Psi_0 = \Sigma + \Gamma_0 \Gamma_0^T$, where Γ_0 was estimated as above. The OLS estimate for F_2 is $\hat{F}_2 = \left(L_0^T \Psi_0^T L_0\right)^{-1} L_0^T \Psi_0^{-1} Z_2 = H_{2_0}^{-1} H_{1_0}$. We can then let U_2 be such that $U_2 \begin{bmatrix} F^T \\ \hat{F}_2^T \end{bmatrix} = 0$ and $\tilde{Z} = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix} U_2$. The starting point for Γ is then

$$\text{2)} \quad \Gamma_0 = \frac{1}{\sqrt{m_1 + m_2 - K}} U_{\tilde{Z}} D_{\tilde{Z}}$$

1.9 Another Formulation of the Problem

Since we do not care about the estimate for L, we would like to be able to enjoy the same luxuries as we do when we estimate the latent factors W from the data. The nice part about that is we get to choose the latent dimension. In order to do that I will introduce a matrix $T \in \mathbb{R}^{s \times K}$, where $s < K$ (if $s = K$ the below procedure reduces to exactly what we have above). The model is then

$$Z_1 = L_{p \times s} T F + \Gamma W_1 + E_1$$

$$Z_2 = L_{p \times s} \underbrace{V}_{MN_{s \times m_2}(0, \Lambda, I_{m_2})} + \Gamma W_2 + E_2$$

All updates remain exactly as above (including the update for $\Lambda_{s \times s}$). The only thing we need change is $A = Z_1 - LTF$. The update for T is then

$$T = \left(L^T \Psi^{-1} L\right)^{-1} L^T \Psi^{-1} Z_1 F^T \left(FF^T\right)^{-1} = H_2^{-1} H_7 F^T \left(FF^T\right)^{-1}.$$

The update for L is

$$L = MU^{-1}$$

$$M = Z_1 F^T T^T + S_2 J_t^{-1} L_t \Lambda = Z_1 F^T T^T + Z_2 G_1^T \Lambda$$

$$U = TFF^T T^T + \Lambda L_t^T J_t^{-1} S_2 J_t^{-1} L_t \Lambda + m_2 \left\{\Lambda - \Lambda L_t^T J_t^{-1} L_t \Lambda\right\} = TFF^T T^T + \Lambda G_1 G_1^T \Lambda + m_2 \left\{\Lambda - \Lambda G_2 \Lambda\right\}$$

Since T and L are not identifiable, at the last iteration I will I will take the QR decomposition of T^T and let $T = Q^T$ and $L = LR^T$.

We can get a possible starting point for T as follows:

1. Model the first set of individuals as $Z_1 = LTF + \Gamma W_1 + E_1 = \tilde{\Gamma}\tilde{W} + E_1$. Determine the latent dimension r_0 using BCV.

2. Rotate out F and re-estimate the latent dimension r_1 using BCV. Set $s = \max(1, r_0 - r_1)$.

3. Estimate $\mu = \tilde{\Gamma}\tilde{W}$ using CATE's factor analysis software.

4. Note that W_1 is independent of F. Therefore, estimate LTF as $\hat{\mu}P_F$.

5. Let $U\Sigma V^T = \widehat{LTF}$ and $\hat{T} = V^T F^T \left(FF^T\right)^{-1}$.

This procedure seems to works well. Another method would be to find the eigenvectors corresponding to the s largest eigenvalues of $\tilde{\Gamma}\tilde{\Gamma}^T - \Gamma\Gamma^T = \tilde{L}\tilde{L}^T$. For example, if $s = 1$ and u was the top eigenvector of $\tilde{L}\tilde{L}^T$, a starting point for T would then be $T^T = \frac{(u^T \Psi^{-1} u)^{-1} u^T \Psi^{-1} Z_1 F^T (FF^T)^{-1}}{\left\|(u^T \Psi^{-1} u)^{-1} u^T \Psi^{-1} Z_1 F^T (FF^T)^{-1}\right\|_2}$ and $\hat{L} = \left\|\left(u^T \Psi^{-1} u\right)^{-1} u^T \Psi^{-1} Z_1 F^T \left(FF^T\right)^{-1}\right\|_2 u$.

1.10   Correcting For Cell Type

Using the equations in the first sections we first rotate out C:

$$\tilde{Y}_1 = B \underbrace{\tilde{X}_1}_{X_1 Q_C} + \Gamma\alpha\tilde{X}_1 + \Gamma \underbrace{\tilde{W}_1}_{W_1 Q_C} + \underbrace{\tilde{E}_1}_{E_1 Q_C}$$

and then compute the QR decomposition of $\tilde{X}_1 = \tilde{R}_1^T \tilde{Q}_1^T$:

$$\tilde{Z}_1 = B\tilde{R}_1^T + \Gamma\left(\alpha\tilde{R}_1^T + \tilde{W}_1^{(1)}\right) + \tilde{E}_1^{(1)}.$$

Note that for $\tilde{R}_1$ small (i.e. cell type explains a large portion of the variability in X_1), we need to account for spurious correlations due to $\tilde{W}_1^{(1)}$ (even if $\alpha = 0$). For the next set of individuals, we have

$$\tilde{Z}_2 = BR_2^T + \Gamma\left(\alpha R_2^T + W_2^{(1)}\right) + L\left(\Omega R_2^T + D\Xi^{(1)}\right) + E_2^{(1)}.$$

For $\Lambda \otimes \left(R_2^{-1} R_2^{-T}\right) = \Lambda \otimes \left(X_2 X_2^T\right)^{-1}$ small in comparison to Ω, we can ignore the spurious correlation caused by $\Xi^{(1)}$. This basically means that if the cellular variability Λ is LESS than the correlation between C and X (like it is in the cases we are considering), then we can ignore spurious correlations, even for small sample sizes in X_2. We can check this by comparing the estimates for D and ΩR_2^T. If we know Ω and D, then we can approximate the distribution for $\Omega + D\Xi^{(1)} R_2^{-T}$. If we were to then estimate $\Omega + D\Xi^{(1)} R_2^{-T}$ directly from the data and found that the estimate strayed too far from Ω, we would know that the data are not informative enough to estimate cell type. If ΩR_2^T is small in

comparison to D, then we should directly estimate $\Omega + D\Xi^{(1)}R_2^{-T}$ from the data.

This idea is important because in some cases we are betting off estimating the confounding directly from the data. I am still working on a method to determine when we should use training data, and the variance of $D\Xi^{(1)}R_2^{-T}$ in relation to Ω will be an important factor.

As of now, I have a method that seems to control the false discovery rate at a nominal level, even for large confounding and noisy estimates of L. There are two scenarios we should consider

1. $\Lambda \otimes \left(X_2 X_2^T\right)^{-1}$ is small. In this case, we can estimate $\Omega + D\Xi^{(1)}R_2^{-T}$ from the data and see how far the estimate is from Ω. If it is close, we know that we can probably estimate the confounding from the data. If not, then we need to use training data. This may be useful in experimental design. If one was involved in a large study (i.e. COPSAC), we could get cell type data on only a fraction of the individuals, estimate L, Ω and $\Lambda = DD^T$. If we find that our estimate for $\Omega + D\Xi^{(1)}R^{-T}$ is close to Ω, then we know the data are informative enough to estimate cell type. If not, then we should collect enough training data to apply my method.

2. $\Lambda \otimes \left(X_2 X_2^T\right)^{-1}$ is large. In this case, you should use the first set of samples with cell type information to estimate L and see if we can reliably estimate cell type composition from the data. If we can, then we can ignore cell type and estimate it from the data.

2   Model 2

In this formulation of the problem, I add the term T so that the mean cell type response is no longer $L_{p\times K}C_{K\times n}$, but $L_{p\times s}T_{s\times K}C_{K\times n}$. T is meant to act as a projection matrix that takes only the elements in C relevant to the mean response. The reason for adding this term came from the observation that if I ignored C in Michelle's data, I estimated the latent dimension to be 4, while when I corrected for C I estimated the latent dimension to be 3. This means that out of the 5 measured cell types, there is a one dimensional projection of C that describes the majority of the relationship between Y and C. The full model is

$$Y_1 = BX_1 + L_{p\times s}T_{s\times K}C_{1_{K\times n_1}} + \underbrace{\Gamma_{p\times r}W_{1_{r\times n_1}}}_{\text{additional confounding}} + \underbrace{E_1}_{MN_{p\times n_1}\left(0,\Sigma_{p\times p},I_{n_1}\right)}$$

$$Y_2 = BX_2 + L_{p\times s}T_{s\times K}C_{2_{K\times n_2}} + \Gamma_{p\times r}W_{2_{r\times n_2}} + E_2$$

We assume here that

$$C_2 = \Omega_{K\times d}X_2 + \underbrace{\Xi}_{MN_{K\times n_2}\left(0,\Lambda_{K\times K},I_{n_2}\right)}$$

and that $W_i \sim MN_{r\times n_i}\left(0, I_r, I_{n_i}\right)$. We use the training date in (C_1, X_1) to estimate Ω and Λ. From now on, I will let $C = C_1$ and write out the second model as

$$Y_2 = BX_2 + LT\Omega X_2 + LT\Xi + \Gamma W_2 + E_2.$$

If we let $Q_{X_i}^T$ be an orthogonal basis for the kernel of X_i and $Z_i = Y_i Q_{X_i}$, we estimate L, Γ and T from Z_1 and Z_2:

$$Z_1 = LT\underbrace{F}_{=CQ_{X_1}} + \Gamma\underbrace{\tilde{W}_1}_{MN_{r\times(n_1-d)}\left(0,I_r,I_{n_1-d}\right)} + \tilde{E}_1$$

$$Z_2 = L\underbrace{V}_{MN_{s\times(n_2-d)}\left(0,T^T\Lambda T,I_{n_2-d}\right)} + \Gamma\tilde{W}_2 + \tilde{E}_2$$

Note that if $s = K$, then we can let $T = I_K$ and the above model reduces to the first case. Notice that for $s < K$, we are now estimating LT, which is a rank deficient matrix, meaning the optimization for L is non-convex. Therefore, in order to estimate $L_{p\times s}$ and $T_{s\times K}$ in this problem, we need a good starting point for $L_{p\times s}$, i.e. we need more training data (actually, whenever $\Gamma \neq 0$, estimating L becomes a challenge when we don't have a lot of training data). The benefit of including the term T is that estimating Γ and L simultaneously is significantly easier, as we do not have to worry about spurious signal coming from the components of C that do not contribute anything to the variability of Y.

2.1 Update for L

$$L = MU^{-1}$$

$$M = Z_1 F^T T^T + S_2 J_t^{-1} L_t \tilde{\Lambda} = Z_1 F^T T^T + Z_2 G_1^T \tilde{\Lambda}$$

$$U = TFF^T T^T + \tilde{\Lambda} L_t^T J_t^{-1} S_2 J_t^{-1} L_t \tilde{\Lambda} + m_2 \left\{ \tilde{\Lambda} - \tilde{\Lambda} L_t^T J_t^{-1} L_t \tilde{\Lambda} \right\} = TFF^T T^T + \tilde{\Lambda} G_1 G_1^T \tilde{\Lambda} + m_2 \left\{ \tilde{\Lambda} - \tilde{\Lambda} G_2 \tilde{\Lambda} \right\}$$

2.2 Update for T

Let $\tilde{\Lambda} = T \Lambda T^T$. This will replace Λ in all of the G matrices and the log-likelihood above. We only use Λ when we update T.

$$\log f(Z_1) = -\frac{m_1}{2} \log|\Psi| - \frac{1}{2} \operatorname{Tr}\left[(Z_1 - LTF)^T \Psi^{-1} (Z_1 - LTF) \right] = Const + \operatorname{Tr}\left(T^T L^T \Psi^{-1} Z_1 F^T \right) - \frac{1}{2} \operatorname{Tr}\left(T^T L^T \Psi^{-1} LTFF^T \right)$$

$$\log f(Z_2) = -\frac{m_2}{2} \log\left| \Psi + LT\Lambda T^T L^T \right| - \frac{1}{2} \operatorname{Tr}\left[S_2 \left(\Psi + LT\Lambda T^T L^T \right)^{-1} \right]$$

If we let $g(T) = -\log f(Z_1) - \log f(Z_2)$, then

$$\frac{dg}{dT} = m_2 L^T J^{-1} LT\Lambda - L^T J^{-1} Z_2 Z_2^T J^{-1} LT\Lambda + L^T \Psi^{-1} LTFF^T - L^T \Psi^{-1} Z_1 F^T = m_2 G_2 T\Lambda - G_1 G_1^T T\Lambda + H_2 TFF^T - H_7 F^T$$

and

$$g(T) = \frac{m_2}{2} \left[\log|\tilde{\Lambda}| + \log|\tilde{\Lambda}^{-1} + H_2| \right] - \operatorname{Tr}\left(H_7 F^T T^T \right) + \frac{1}{2} \operatorname{Tr}\left(H_2 TFF^T T^T \right) - \frac{1}{2} \operatorname{Tr}\left[H_1^T \left(\tilde{\Lambda}^{-1} + H_2 \right)^{-1} H_1 \right].$$

I can use g and $\frac{dg}{dT}$ as input into BFGS.

I will perform these updates as follows:
When I update L, I will update H_1, H_2, H_7. I will then update T. Once I have T, I will update $\tilde{\Lambda}, A, H_3, H_5$, all G's.

2.3 Update for Γ

$$\Gamma = MU^{-1}$$

$$M = AA^T \Psi_t^{-1} \Gamma_t + S_2 J_t^{-1} \Gamma_t - L \left\{ \tilde{\Lambda} L^T J_t^{-1} S_2 J_t^{-1} \Gamma_t - m_2 \tilde{\Lambda} L^T J_t^{-1} \Gamma_t \right\} = AG_3 + Z_2 G_4 - L \left\{ \tilde{\Lambda} G_1 G_4 - m_2 \tilde{\Lambda} G_5 \right\}$$

$$U = \left(\Gamma_t^T \Psi_t^{-1} AA^T \Psi_t^{-1} \Gamma_t + m_1 \left\{ I_r - \Gamma_t^T \Psi_t^{-1} \Gamma_t \right\} \right) + \left(\Gamma_t^T J_t^{-1} S_2 J_t^{-1} \Gamma_t + m_2 \left\{ I_r - \Gamma_t^T J_t^{-1} \Gamma_t \right\} \right) = G_3^T G_3 + m_1 \left\{ I_r - G_6 \right\} + G_4^T G_4 + m_2 \left\{ I_r - G_7 \right\}.$$

I then need to update $H_1 - H_7$ and then $G_1 - G_7$.

2.4 Update for Σ

$$(m_1 + m_2) v = \operatorname{diag}\left(AA^T \right) + \operatorname{diag}(S_2) - 2 \operatorname{diag}\left(AA^T \Psi^{-1} \Gamma \Gamma^T \right) - 2 \operatorname{diag}\left(S_2 J^{-1} \Gamma \Gamma^T \right) + \operatorname{diag}\left[\Gamma \left(\Gamma^T \Psi^{-1} AA^T \Psi^{-1} \Gamma + m_1 \left\{ I_r - \Gamma^T \Psi^{-1} \Gamma \right\} \right) \Gamma^T \right]$$

$$+ 2 \operatorname{diag}\left[L \left(\tilde{\Lambda} L^T J^{-1} S_2 J^{-1} \Gamma - m_2 \tilde{\Lambda} L^T J^{-1} \Gamma \right) \Gamma^T \right] + \operatorname{diag}\left[\Gamma \left(\Gamma^T J^{-1} S_2 J^{-1} \Gamma + m_2 \left\{ I_r - \Gamma^T J^{-1} \Gamma \right\} \right) \Gamma^T \right]$$

$$+ \operatorname{diag}\left[L \left(\tilde{\Lambda} L^T J^{-1} S_2 J^{-1} L\tilde{\Lambda} + m_2 \left\{ \tilde{\Lambda} - \tilde{\Lambda} L^T J^{-1} L\tilde{\Lambda} \right\} \right) L^T \right] - 2 \operatorname{diag}\left(S_2 J^{-1} L\tilde{\Lambda} L^T \right)$$

$$= \operatorname{diag}\left(AA^T \right) + \operatorname{diag}(S_2) + \operatorname{diag}\left[\left(\Gamma \left(G_3^T G_3 + m_1 \left\{ I_r - G_6 \right\} + G_4^T G_4 + m_2 \left\{ I_r - G_7 \right\} \right) + 2L \left(\tilde{\Lambda} G_1 G_4 - m_2 \tilde{\Lambda} G_5 \right) - 2AG_3 - 2Z_2 G_4 \right) \Gamma^T \right]$$

$$+ \operatorname{diag}\left[\left(L \left(\tilde{\Lambda} G_1 G_1^T \tilde{\Lambda} + m_2 \left\{ \tilde{\Lambda} - \tilde{\Lambda} G_2 \tilde{\Lambda} \right\} \right) - 2Z_2 G_1^T \tilde{\Lambda} \right) L^T \right].$$

I then need to update $H_1 - H_6$ and $G_1 - G_7$.

The above simulations indicate that under this scenario, we do poorly. For p very large (and r small), this appears to happen when $\frac{1}{p} \Gamma^T \Sigma^{-1} \Gamma = O\left(\frac{1}{n} \right)$ and the correlation $\frac{1}{p} \Gamma^T \Sigma^{-1} YX \left(X^T X \right)^{-1}$ is somewhat large. Note that as n gets larger, the correlation we would need for this method to fail also becomes larger (for fixed r). This is because:

$$\hat{\Sigma}^{-1/2} \hat{\Gamma} = \hat{\Sigma}^{-1/2} \Gamma + R, \quad \text{where} \quad \frac{1}{p} R^T R \approx \frac{1}{n} I_r \text{ for } p \text{ large}.$$

$$\Rightarrow \frac{1}{p}\hat{\Gamma}^T\hat{\Sigma}^{-1}\hat{\Gamma} = \frac{1}{p}\Gamma^T\hat{\Sigma}^{-1}\Gamma + \frac{1}{p}R^TR + \underbrace{\frac{1}{p}R^T\hat{\Sigma}^{-1/2}\Gamma + \frac{1}{p}\Gamma^T\hat{\Sigma}^{-1/2}R}_{\approx 0} \approx \frac{1}{p}\Gamma^T\hat{\Sigma}^{-1}\Gamma + \frac{1}{n}I_r \approx \frac{1}{p}\Gamma^T\Sigma^{-1}\Gamma + \frac{1}{n}I_r$$