ST 762, HOMEWORK 6 EXTRA PROBLEMS SOLUTIONS, FALL 2009

1. (a) There are lots of ways to do this; here, we just do it by "brute force."

From the assumed approximate model, we know that $E(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta}) = \boldsymbol{0}$ and thus $E(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta})^T = \boldsymbol{D} + \boldsymbol{C}_i$. We also have that $E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \boldsymbol{0}$ and thus

$$E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T = \left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\left(\sum_{k=1}^m \boldsymbol{A}_k^T(\boldsymbol{D}+\boldsymbol{C}_k)\boldsymbol{A}_k\right)\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}.$$

We also have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\sum_{k=1}^m \boldsymbol{A}_k^T(\hat{\boldsymbol{\beta}}_k - \boldsymbol{A}_k\boldsymbol{\beta}).$$

Consider a summand in $\hat{\boldsymbol{D}}_{STS}$. We can write

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\hat{\boldsymbol{\beta}})^T &= E(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta})^T \\
&+ E(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta})(\boldsymbol{A}_i\boldsymbol{\beta} - \boldsymbol{A}_i\hat{\boldsymbol{\beta}})^T \\
&+ E(\boldsymbol{A}_i\boldsymbol{\beta} - \boldsymbol{A}_i\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta})^T \\
&+ \boldsymbol{A}_i E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\boldsymbol{A}_i^T.
\end{aligned}$$

The first term on the right hand side is just $\boldsymbol{D} + \boldsymbol{C}_i$, and the last term is from above

$$\boldsymbol{A}_i\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\left(\sum_{k=1}^m \boldsymbol{A}_k^T(\boldsymbol{D}+\boldsymbol{C}_k)\boldsymbol{A}_k\right)\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\boldsymbol{A}_i^T.$$

Consider the third term; the second is just its transpose. This term may be written

$$-\boldsymbol{A}_i\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\sum_{k=1}^m \boldsymbol{A}_k^T(\hat{\boldsymbol{\beta}}_k - \boldsymbol{A}_k\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta})^T.$$

Now by assumption, the $\hat{\boldsymbol{\beta}}_i$ are independent, so have 0 covariance. Thus, this term becomes

$$-\boldsymbol{A}_i\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\boldsymbol{A}_i^T(\boldsymbol{D}+\boldsymbol{C}_i),$$

and the second is the transpose of this term. So, placing in the sum and combining, we get

$$\boldsymbol{D} + (m-1)^{-1}(\boldsymbol{D} + \sum_{i=1}^m \boldsymbol{C}_i)$$

$$+(m-1)^{-1}\sum_{i=1}^m \boldsymbol{A}_i\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\left(\sum_{k=1}^m \boldsymbol{A}_k^T(\boldsymbol{D}+\boldsymbol{C}_k)\boldsymbol{A}_k\right)\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\boldsymbol{A}_i^T$$

$$-(m-1)^{-1}\sum_{i=1}^m \boldsymbol{A}_i\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\boldsymbol{A}_i^T(\boldsymbol{D}+\boldsymbol{C}_i)$$

$$-(m-1)^{-1}\sum_{i=1}^m(\boldsymbol{D}+\boldsymbol{C}_i)\boldsymbol{A}_i\left(\sum_{k=1}^m \boldsymbol{A}_k^T\boldsymbol{A}_k\right)^{-1}\boldsymbol{A}_i^T.$$

One could manipulate this further, but it is fairly clear that the terms after the first do not equal 0 in general. Thus, $\hat{\boldsymbol{D}}_{STS}$ is likely not unbiased in general.

(b) We now have that

$$\left(\sum_{k=1}^{m} \boldsymbol{A}_k^T \boldsymbol{A}_k\right)^{-1} = m^{-1}\boldsymbol{I}_b.$$

Inserting this and $\boldsymbol{A}_i = \boldsymbol{I}_b$ into the above, we get

$$\boldsymbol{D} + \boldsymbol{B},$$

where

$$\boldsymbol{B} = \{(m-1)^{-1} + (m-1)^{-1} - (m-1)^{-1} - (m-1)^{-1}\}\boldsymbol{D}$$
$$+\{(m-1)^{-1} + (m-1)^{-1}m^{-1} - (m-1)^{-1}m^{-1} - (m-1)^{-1}m^{-1}\}\sum_{i=1}^{m} \boldsymbol{C}_i.$$

This reduces to $\boldsymbol{B} = m^{-1}\sum_{i=1}^{m} \boldsymbol{C}_i$, so that

$$E(\hat{\boldsymbol{D}}_{STS}) = \boldsymbol{D} + m^{-1}\sum_{i=1}^{m} \boldsymbol{C}_i.$$

As $\boldsymbol{B}$ is the sum of assumed positive definite covariance matrices, it is reasonable to assume that is nonnegative definite. Thus, we would expect $\hat{\boldsymbol{D}}_{STS}$ to "overestimate" $\boldsymbol{D}$ in the sense of definiteness.

2. (a) Because of the independence and normality assumptions, we can write $\log p(\boldsymbol{U}, \boldsymbol{b}; \boldsymbol{\theta}) = \log p(\boldsymbol{U}|\boldsymbol{b}; \boldsymbol{\theta}) + \log p(\boldsymbol{b}; \boldsymbol{\theta})$ as, ignoring constants,

$$-\frac{1}{2}\sum_{i=1}^{m}\left\{(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \boldsymbol{b}_i)^T \boldsymbol{C}_i^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \boldsymbol{b}_i) - \log|\boldsymbol{D}|\boldsymbol{b}_i^T \boldsymbol{D}^{-1}\boldsymbol{b}_i\right\}. \quad (1)$$

Note from the assumptions that $(\hat{\boldsymbol{\beta}}_i^T, \boldsymbol{b}_i^T)^T$ is multivariate normal with mean $\{\boldsymbol{0}^T, (\boldsymbol{A}_i\boldsymbol{\beta})^T\}^T$ and covariance matrix

$$\begin{pmatrix} \boldsymbol{D} & \boldsymbol{D} \\ \boldsymbol{D} & \boldsymbol{D} + \boldsymbol{C}_i \end{pmatrix}.$$

From standard results for the multivariate normal distribution, it follows that the distribution of $\boldsymbol{b}_i$ given $\hat{\boldsymbol{\beta}}_i$ is $\mathcal{N}\{\boldsymbol{D}(\boldsymbol{D}+\boldsymbol{C}_i)^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta}), \boldsymbol{D} - \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{C}_i)^{-1}\boldsymbol{D}\}$. Thus, $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_{(k)})$ is the expectation of (1) with respect to this distribution with $\boldsymbol{\beta}$ and $\boldsymbol{D}$ replaced by $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{D}}$. Let $\hat{\boldsymbol{b}}_i = \hat{\boldsymbol{D}}_{(k)}(\hat{\boldsymbol{D}}_{(k)} + \boldsymbol{C}_i)^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta}_{(k)})$, and note that $\hat{\boldsymbol{M}}_i = \hat{\boldsymbol{D}}_{(k)} - \hat{\boldsymbol{D}_{(k)}}(\hat{\boldsymbol{D}}_{(k)} + \boldsymbol{C}_i)^{-1}\hat{\boldsymbol{D}}_{(k)}\} = (\hat{\boldsymbol{D}_{(k)}}^{-1} + \boldsymbol{C}_i^{-1})^{-1}$. Then, with $\boldsymbol{Y} = \hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \boldsymbol{b}_i)$, the conditional distribution of $\boldsymbol{Y}$ given $\hat{\boldsymbol{\beta}}_i$ evaluated at $\hat{\boldsymbol{\beta}}_{(k)}$ and $\hat{\boldsymbol{D}}_{(k)}$ is $\mathcal{N}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \hat{\boldsymbol{b}}_i, \hat{\boldsymbol{M}}_i)$. Thus, the expectation of the first term in (1) with respect to this distribution is, using the hint with $\boldsymbol{Y}$ as above and $\boldsymbol{A} = \boldsymbol{C}_i^{-1}$,

$$(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \hat{\boldsymbol{b}}_i)^T \boldsymbol{C}_i^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \hat{\boldsymbol{b}}_i) + \mathrm{tr}(\boldsymbol{C}_i^{-1}\hat{\boldsymbol{M}}_i).$$

Similarly, the expectation of the third term in (1) is

$$\hat{\boldsymbol{b}}_i^T \boldsymbol{D}^{-1}\hat{\boldsymbol{b}}_i + \mathrm{tr}(\boldsymbol{D}^{-1}\hat{\boldsymbol{M}}_i).$$

Thus, putting it all together, we have

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_{(k)}) &= -\frac{1}{2}\sum_{i=1}^{m}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \hat{\boldsymbol{b}}_i)^T \boldsymbol{C}_i^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \hat{\boldsymbol{b}}_i) + -\frac{1}{2}\sum_{i=1}^{m}\mathrm{tr}(\boldsymbol{C}_i^{-1}\hat{\boldsymbol{M}}_i) \\ &\quad -\frac{m}{2}\log|\boldsymbol{D}| - \frac{1}{2}\sum_{i=1}^{m}\hat{\boldsymbol{b}}_i^T\boldsymbol{D}^{-1}\hat{\boldsymbol{b}}_i - \frac{1}{2}\sum_{i=1}^{m}\mathrm{tr}(\boldsymbol{D}^{-1}\hat{\boldsymbol{M}}_i) \end{aligned} \tag{2}$$

Using a standard result for quadratic forms, $\hat{\boldsymbol{b}}_i^T\boldsymbol{D}^{-1}\hat{\boldsymbol{b}}_i = \mathrm{tr}\{\boldsymbol{D}^{-1}(\hat{\boldsymbol{b}}_i\hat{\boldsymbol{b}}_i^T)\}$, we may rewrite (2) as

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_{(k)}) &= -\frac{1}{2}\sum_{i=1}^{m}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \hat{\boldsymbol{b}}_i)^T \boldsymbol{C}_i^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta} - \hat{\boldsymbol{b}}_i) + -\frac{1}{2}\sum_{i=1}^{m}\mathrm{tr}(\boldsymbol{C}_i^{-1}\hat{\boldsymbol{M}}_i) \\ &\quad -\frac{m}{2}\log|\boldsymbol{D}| - \frac{1}{2}\sum_{i=1}^{m}\mathrm{tr}\{\boldsymbol{D}^{-1}(\hat{\boldsymbol{b}}_i\hat{\boldsymbol{b}}_i^T + \hat{\boldsymbol{M}}_i)\}, \end{aligned} \tag{3}$$

which will come in handy later.

(b) We have

$$\boldsymbol{A}_i\hat{\boldsymbol{\beta}}_{(k)} + \hat{\boldsymbol{b}}_i = \boldsymbol{A}_i\hat{\boldsymbol{\beta}}_{(k)} + \hat{\boldsymbol{D}}_{(k)}(\hat{\boldsymbol{D}}_{(k)} + \boldsymbol{C}_i)^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{A}_i\boldsymbol{\beta}_{(k)}).$$

We may rewrite this as

$$\{\boldsymbol{I} - \hat{\boldsymbol{D}}_{(k)}(\hat{\boldsymbol{D}}_{(k)} + \boldsymbol{C}_i)^{-1}\}\boldsymbol{A}_i\hat{\boldsymbol{\beta}}_{(k)} + \hat{\boldsymbol{D}}_{(k)}(\hat{\boldsymbol{D}}_{(k)} + \boldsymbol{C}_i)^{-1}\hat{\boldsymbol{\beta}}_i.$$

For the first term, recalling $(\boldsymbol{D} + \boldsymbol{C}_i)^{-1} = \boldsymbol{D}^{-1}(\boldsymbol{D}^{-1} + \boldsymbol{C}_i^{-1})^{-1}\boldsymbol{C}_i^{-1}$, we have that the first term is

$$(\hat{\boldsymbol{D}}_{(k)}^{-1} + \boldsymbol{C}_i^{-1})^{-1}(\hat{\boldsymbol{D}}_{(k)}^{-1} + \boldsymbol{C}_i^{-1} - \boldsymbol{C}_i^{-1})\boldsymbol{A}_i\hat{\boldsymbol{\beta}}_{(k)} = (\hat{\boldsymbol{D}}_{(k)}^{-1} + \boldsymbol{C}_i^{-1})^{-1}\hat{\boldsymbol{D}}_{(k)}^{-1}\boldsymbol{A}_i\hat{\boldsymbol{\beta}}_{(k)}.$$

The second term similarly equals

$$(\hat{\boldsymbol{D}}_{(k)}^{-1} + \boldsymbol{C}_i^{-1})^{-1}\boldsymbol{C}_i^{-1}\hat{\boldsymbol{\beta}}_i,$$

and putting them together yields the result.

(c) As $\hat{\boldsymbol{\beta}}_{(k+1)} = \hat{\boldsymbol{\beta}}_{(k)}$, just write $\hat{\boldsymbol{\beta}}$, and just write $\boldsymbol{D}$ for simplicity, as $\hat{\boldsymbol{D}}_{(k)}$ is just fixed here. From page 432, substituting for $\tilde{\boldsymbol{\beta}}_{i,(k+1)}$ and using $(\boldsymbol{D} + \boldsymbol{C}_i)^{-1} = \boldsymbol{D}^{-1}(\boldsymbol{D}^{-1} + \boldsymbol{C}^{-1})^{-1}\boldsymbol{C}_i^{-1}$, we have

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{m}\boldsymbol{A}_i^T\boldsymbol{D}^{-1}\boldsymbol{A}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{A}_i^T\boldsymbol{D}^{-1}(\boldsymbol{D}^{-1} + \boldsymbol{C}^{-1})^{-1}(\boldsymbol{D}^{-1}\boldsymbol{A}_i\hat{\boldsymbol{\beta}} + \boldsymbol{C}_i^{-1}\hat{\boldsymbol{\beta}}_i) \\ &= \left(\sum_{i=1}^{m}\boldsymbol{A}_i^T\boldsymbol{D}^{-1}\boldsymbol{A}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{A}_i^T(\boldsymbol{D} + \boldsymbol{C}_i)^{-1}\hat{\boldsymbol{\beta}}_i \\ &\quad + \left(\sum_{i=1}^{m}\boldsymbol{A}_i^T\boldsymbol{D}^{-1}\boldsymbol{A}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{A}_i^T\boldsymbol{D}^{-1}(\boldsymbol{D}^{-1} + \boldsymbol{C}^{-1})^{-1}\boldsymbol{D}^{-1}\boldsymbol{A}_i\hat{\boldsymbol{\beta}}. \end{aligned}$$

We may thus rearrange to obtain

$$\left(\sum_{i=1}^{m}\boldsymbol{A}_i^T\boldsymbol{D}^{-1}\boldsymbol{A}_i\right)^{-1}\left\{\sum_{i=1}^{m}\boldsymbol{A}_i^T\boldsymbol{D}^{-1}\boldsymbol{A}_i - \sum_{i=1}^{m}\boldsymbol{A}_i^T\boldsymbol{D}^{-1}(\boldsymbol{D}^{-1} + \boldsymbol{C}^{-1})^{-1}\boldsymbol{D}^{-1}\boldsymbol{A}_i\right\}\hat{\boldsymbol{\beta}}$$

$$= \left(\sum_{i=1}^{m} \boldsymbol{A}_i^T \boldsymbol{D}^{-1} \boldsymbol{A}_i \right)^{-1} \sum_{i=1}^{m} \boldsymbol{A}_i^T (\boldsymbol{D} + \boldsymbol{C}_i)^{-1} \hat{\boldsymbol{\beta}}_i.$$

This may be simplified to

$$\left\{ \sum_{i=1}^{m} \boldsymbol{A}_i^T \left(\boldsymbol{D}^{-1} - \boldsymbol{D}^{-1} (\boldsymbol{D}^{-1} + \boldsymbol{C}^{-1})^{-1} \boldsymbol{D}^{-1} \right) \boldsymbol{A}_i \right\} \hat{\boldsymbol{\beta}} = \sum_{i=1}^{m} \boldsymbol{A}_i^T (\boldsymbol{D} + \boldsymbol{C}_i)^{-1} \hat{\boldsymbol{\beta}}_i.$$

As $\boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}(\boldsymbol{D}^{-1} + \boldsymbol{C}^{-1})^{-1}\boldsymbol{D}^{-1} = (\boldsymbol{D} + \boldsymbol{C}_i)^{-1}$, the result follows.

(d) From (a), we want to maximize in $\boldsymbol{D}$

$$\frac{m}{2} \log |\boldsymbol{D}| - \frac{1}{2} \sum_{i=1}^{m} \mathrm{tr}\{\boldsymbol{D}^{-1} (\hat{\boldsymbol{b}}_i \hat{\boldsymbol{b}}_i^T + \hat{\boldsymbol{M}}_i)\}.$$

We may rewrite this as

$$-\frac{m}{2} \left[\log |\boldsymbol{D}| + \mathrm{tr} \left\{ \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{D}^{-1} (\hat{\boldsymbol{b}}_i \hat{\boldsymbol{b}}_i^T + \hat{\boldsymbol{M}}_i) \right\} \right].$$

Thus, we want to minimize

$$\log |\boldsymbol{D}| + \mathrm{tr} \left\{ \boldsymbol{D}^{-1} \frac{1}{m} \sum_{i=1}^{m} (\hat{\boldsymbol{b}}_i \hat{\boldsymbol{b}}_i^T + \hat{\boldsymbol{M}}_i) \right\}.$$

Using the hint, we find that the minimizer is thus

$$\hat{\boldsymbol{D}}_{(k+1)} = \frac{1}{m} \sum_{i=1}^{m} (\hat{\boldsymbol{b}}_i \hat{\boldsymbol{b}}_i^T + \hat{\boldsymbol{M}}_i).$$

Substituting the definitions of $\hat{\boldsymbol{b}}_i$ (replacing $\hat{\boldsymbol{\beta}}_{(k+1)}$ by $\hat{\boldsymbol{\beta}}_{(k)}$) and $\hat{\boldsymbol{M}}_i$ from (a) and using (b), the result follows.

This problem thus shows that the "usual" "EM algorithm" reported in the notes is really not a "formal" EM algorithm. A truly formal one would maximize $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_{(k)})$ in both $\boldsymbol{\beta}$ and $\boldsymbol{D}$. Instead, this algorithm uses EM only to maximize what one might think of as the "PL" for $\boldsymbol{D}$ and continues to use "GLS" to estimate $\boldsymbol{\beta}$.

3. We will write $\boldsymbol{R}_i$, $\boldsymbol{Z}_i$, etc for short here, so that

$$\hat{\boldsymbol{b}}_i = \boldsymbol{D} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{f}_i).$$

There are $\hat{\boldsymbol{b}}_i$ in $\boldsymbol{Z}_i$ and $\boldsymbol{f}_i$, but this doesn't matter to the following argument, as all we are trying to do is reexpress (15.51) as (15.53).

There are two things to show. We can simplify (15.51) to be

$$p(\boldsymbol{Y}_i | \boldsymbol{x}_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{D}) \approx (2\pi)^{-n_i/2} |\boldsymbol{R}_i|^{-1/2} |\boldsymbol{D}|^{-1/2} |\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i|^{-1/2}$$
$$\times \exp\{-(1/2)(\boldsymbol{Y}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i (\boldsymbol{Y}_i - \boldsymbol{f}_i) - (1/2) \hat{\boldsymbol{b}}_i^T \boldsymbol{D}^{-1} \hat{\boldsymbol{b}}_i\}.$$

Thus, to show the equivalence, we first want to show that

$$|\boldsymbol{R}_i| |\boldsymbol{D}| |\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i| = |\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i|. \tag{4}$$

Here is a famous general result that is handy for this; this result may be found in most books on matrix algebra: If $\boldsymbol{A}$ is $(p \times q)$ and $\boldsymbol{B}$ is $(q \times p)$, then

$$|\boldsymbol{I}_p + \boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{I}_q + \boldsymbol{B}\boldsymbol{A}|.$$

We apply this result to

$$|\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i| = |\boldsymbol{D}^{-1}(\boldsymbol{I} + \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i)| = |\boldsymbol{D}|^{-1} |\boldsymbol{I} + \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i|.$$

The second term in the last expression is thus equal to

$$|\boldsymbol{I} + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1}| = |\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T||\boldsymbol{R}_i|^{-1}.$$

Putting all together, we have

$$|\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i| = |\boldsymbol{D}|^{-1} |\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T||\boldsymbol{R}_i|^{-1}.$$

Thus, from above, we may conclude the result (4).

Now we need to deal with the term in the exponential. We want to show that

$$(\boldsymbol{Y}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{f}_i) + \hat{\boldsymbol{b}}_i^T \boldsymbol{D}^{-1} \hat{\boldsymbol{b}}_i = (\boldsymbol{u}_i - \boldsymbol{f}_i)(\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T)^{-1}(\boldsymbol{u}_i - \boldsymbol{f}_i), \tag{5}$$

where we have defined

$$\boldsymbol{u}_i = \boldsymbol{Y}_i + \boldsymbol{Z}_i \hat{\boldsymbol{b}}_i$$

so that $(\boldsymbol{u}_i - \boldsymbol{f}_i) = (\boldsymbol{Y}_i - \boldsymbol{h}_i)$ and $\boldsymbol{h}_i$ is defined on page 444. Note that we can write

$$\hat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i - \boldsymbol{Z}_i \hat{\boldsymbol{b}}_i),$$

which leads to

$$\hat{\boldsymbol{b}}_i + \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i \hat{\boldsymbol{b}}_i = \boldsymbol{D}(\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i)\hat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i)$$

so that we finally obtain

$$\hat{\boldsymbol{b}}_i = (\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i).$$

Note that we can write the left hand side of (5) as

$$(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) - \hat{\boldsymbol{b}}_i^T \boldsymbol{Z}_i \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) - (\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i \hat{\boldsymbol{b}}_i + \hat{\boldsymbol{b}}_i^T (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1}) \hat{\boldsymbol{b}}_i.$$

Now simplifying this and inserting the expression above for $\hat{\boldsymbol{b}}_i$, we obtain

$$\begin{gathered}(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) - (\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) \\ -(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) \\ +(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})(\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i).\end{gathered}$$

This simplifies further to

$$(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i) - (\boldsymbol{u}_i - \boldsymbol{f}_i)^T \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1} (\boldsymbol{u}_i - \boldsymbol{f}_i)$$

which can be rewritten as

$$(\boldsymbol{u}_i - \boldsymbol{f}_i)^T \{\boldsymbol{R}_i^{-1} - \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{R}^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{R}_i^{-1}\} (\boldsymbol{u}_i - \boldsymbol{f}_i).$$

Now the middle term can be seen to be equal to (c.f. page 250 of the notes)

$$(\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}\boldsymbol{Z}_i^T)^{-1}.$$

Substituting this gives the desired result.