Expectation maximization of a linear dynamical system with a separate offset for each data set and an emissions matrix with columns that sum to 1

The goal of this derivation is to find the EM update for the C and $\boldsymbol{d}^{(i)}$ parameters of an LDS. Specifically, we want to fit a separate $\boldsymbol{d}^{(i)}$ for each data set. Finally, we also want to restrict C such that all the columns sum to 1.

LDS equations where data sets indexed with i

$\boldsymbol{x}_t^{(i)} = A\boldsymbol{x}_{t-1}^{(i)} + B\boldsymbol{u}_t^{(i)} + \boldsymbol{w}_t^{(i)}$

$\boldsymbol{y}_t^{(i)} = C\boldsymbol{x}_t^{(i)} + D\boldsymbol{u}_t^{(i)} + \boldsymbol{d}^{(i)} + \boldsymbol{v}_t^{(i)}$

$\boldsymbol{w}_t^{(i)} \sim \mathcal{N}(0, Q)$, $\boldsymbol{v}_t^{(i)} \sim \mathcal{N}(0, R)$

Consider the term of the complete data log likelihood L_y which deals with emissions $\boldsymbol{y}$. This is taken from equation 73 in Jonathan's Kalman Filter tutorial. T is the number of time points and W is the number of worms (data sets). All vectors are assumed to be column vectors

$$\begin{gathered} L_y = \mathbb{E}\left[\sum_{t=1}^{T^{(i)}} \log \mathcal{N}(\boldsymbol{y}_t^{(i)}; C\boldsymbol{x}_t^{(i)} + D\boldsymbol{u}_t^{(i)} + \boldsymbol{d}^{(i)}, R)\right] = \\ -\tfrac{1}{2}\sum_{i=1}^{W}\sum_{t=1}^{T} \mathbb{E}\left[(\boldsymbol{y}_t^{(i)} - C\boldsymbol{x}_t^{(i)} - D\boldsymbol{u}_t^{(i)} - \boldsymbol{d}^{(i)})^\intercal R^{-1}(\boldsymbol{y}_t^{(i)} - C\boldsymbol{x}_t^{(i)} - D\boldsymbol{u}_t^{(i)} - \boldsymbol{d}^{(i)})\right] \\ -\tfrac{1}{2}\sum_{i=1}^{W} T^{(i)} \log|2\pi R| \end{gathered} \tag{1}$$

Expand the quadratic into all of its terms

$$L_y = -\frac{1}{2}\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\begin{bmatrix} \boldsymbol{y}_t^{(i)\intercal} R^{-1}\boldsymbol{y}_t^{(i)} \\ -\boldsymbol{y}_t^{(i)\intercal} R^{-1} C\boldsymbol{x}_t^{(i)} - \boldsymbol{x}^{(i)\intercal} C^\intercal R^{-1}\boldsymbol{y}_t^{(i)} \\ -\boldsymbol{y}_t^{(i)\intercal} R^{-1} D\boldsymbol{u}_t^{(i)} - \boldsymbol{u}_t^{(i)\intercal} D^\intercal R^{-1}\boldsymbol{y}_t^{(i)} \\ -\boldsymbol{y}_t^{(i)\intercal} R^{-1}\boldsymbol{d}^{(i)} - \boldsymbol{d}^{(i)\intercal} R^{-1}\boldsymbol{y}_t^{(i)} \\ +\boldsymbol{x}_t^{(i)\intercal} C^\intercal R^{-1} C\boldsymbol{x}_t^{(i)} \\ +\boldsymbol{x}_t^{(i)\intercal} C^\intercal R^{-1} D\boldsymbol{u}_t^{(i)} + \boldsymbol{u}_t^{(i)\intercal} D^\intercal R^{-1} C\boldsymbol{x}_t^{(i)} \\ +\boldsymbol{x}_t^{(i)\intercal} C^\intercal R^{-1}\boldsymbol{d}^{(i)} + \boldsymbol{d}^{(i)\intercal} R^{-1} C\boldsymbol{x}_t^{(i)} \\ +\boldsymbol{u}_t^{(i)\intercal} D^\intercal R^{-1} D\boldsymbol{u}_t^{(i)} \\ +\boldsymbol{u}_t^{(i)\intercal} D^\intercal R^{-1}\boldsymbol{d}^{(i)} + \boldsymbol{d}^{(i)\intercal} R^{-1} D\boldsymbol{u}_t^{(i)} \\ +\boldsymbol{d}^{(i)\intercal} R^{-1}\boldsymbol{d}^{(i)} \end{bmatrix} - \frac{1}{2}\sum_{i=1}^{W} T^{(i)} \log|2\pi R| \tag{2}$$

Because this entire term is a scalar, the entire thing can be written as a trace. Trace distributes across addition and matrix multiplication within a trace can be circularly permuted. Permute all the terms so that R^{-1} is on the left

$$
L_y = -\frac{1}{2}\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\begin{array}{c}
\mathrm{Tr}\left[R^{-1}\boldsymbol{y}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right] \\
-\mathrm{Tr}\left[R^{-1}C\boldsymbol{x}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right] - \mathrm{Tr}\left[R^{-1}\boldsymbol{y}_t^{(i)}\boldsymbol{x}^{(i)\intercal}C^\intercal\right] \\
-\mathrm{Tr}\left[R^{-1}D\boldsymbol{u}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right] - \mathrm{Tr}\left[R^{-1}\boldsymbol{y}_t^{(i)}\boldsymbol{u}_t^{(i)\intercal}D^\intercal\right] \\
-\mathrm{Tr}\left[R^{-1}\boldsymbol{d}^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right] - \mathrm{Tr}\left[R^{-1}\boldsymbol{y}_t^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}C\boldsymbol{x}_t^{(i)}\boldsymbol{x}_t^{(i)\intercal}C^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}D\boldsymbol{u}_t^{(i)}\boldsymbol{x}_t^{(i)\intercal}C^\intercal\right] + \mathrm{Tr}\left[R^{-1}C\boldsymbol{x}_t^{(i)}\boldsymbol{u}_t^{(i)\intercal}D^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}\boldsymbol{d}^{(i)}\boldsymbol{x}_t^{(i)\intercal}C^\intercal\right] + \mathrm{Tr}\left[R^{-1}C\boldsymbol{x}_t^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}D\boldsymbol{u}_t^{(i)}\boldsymbol{u}_t^{(i)\intercal}D^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}\boldsymbol{d}^{(i)}\boldsymbol{u}_t^{(i)\intercal}D^\intercal\right] + \mathrm{Tr}\left[R^{-1}D\boldsymbol{u}_t^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}\boldsymbol{d}^{(i)}\boldsymbol{d}^{(i)\intercal}\right]
\end{array}\right] - \frac{1}{2}\sum_{i=1}^{W}T^{(i)}\log|2\pi R| \tag{3}
$$

Distribute the expectation and the sums

$$
L_y = -\frac{1}{2}\left[\begin{array}{c}
\mathrm{Tr}\left[R^{-1}\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{y}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right]\right] \\
-\mathrm{Tr}\left[R^{-1}C\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{x}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right]\right] - \mathrm{Tr}\left[R^{-1}\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{y}_t^{(i)}\boldsymbol{x}^{(i)\intercal}\right]C^\intercal\right] \\
-\mathrm{Tr}\left[R^{-1}D\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right]\right] - \mathrm{Tr}\left[R^{-1}\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{y}_t^{(i)}\boldsymbol{u}_t^{(i)\intercal}\right]D^\intercal\right] \\
-\mathrm{Tr}\left[R^{-1}\sum_{i=1}^{W}\boldsymbol{d}^{(i)}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{y}_t^{(i)\intercal}\right]\right] - \mathrm{Tr}\left[R^{-1}\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{y}_t^{(i)}\right]\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}C\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{x}_t^{(i)}\boldsymbol{x}_t^{(i)\intercal}\right]C^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}D\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)}\boldsymbol{x}_t^{(i)\intercal}\right]C^\intercal\right] + \mathrm{Tr}\left[R^{-1}C\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{x}_t^{(i)}\boldsymbol{u}_t^{(i)\intercal}\right]D^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^{W}\boldsymbol{d}^{(i)}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{x}_t^{(i)\intercal}\right]C^\intercal\right] + \mathrm{Tr}\left[R^{-1}C\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{x}_t^{(i)}\right]\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}D\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)}\boldsymbol{u}_t^{(i)\intercal}\right]D^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^{W}\boldsymbol{d}^{(i)}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)\intercal}\right]D^\intercal\right] + \mathrm{Tr}\left[R^{-1}D\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)}\right]\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{d}^{(i)}\boldsymbol{d}^{(i)\intercal}\right]\right]
\end{array}\right] - \frac{1}{2}\sum_{i=1}^{W}T^{(i)}\log|2\pi R| \tag{4}
$$

apply the expectation and calculate the following suff stats. Note that Y and $\tilde{Y}$ get modified by imputing missing y values, though no other stats are affected. I skip that here, but adjust in the code.

$$
\begin{array}{rrr}
Y = \sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{y}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right] & \tilde{Y} = \sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{x}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right] & \bar{Y}^{(i)} = \sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{y}_t^{(i)\intercal}\right] \\
M_{(1,T)} = \sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{x}_t^{(i)}\boldsymbol{x}_t^{(i)\intercal}\right] & U_y = \sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)}\boldsymbol{y}_t^{(i)\intercal}\right] & \bar{M}^{(i)\intercal} = \sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{x}_t^{(i)\intercal}\right] \\
U_{(1,T)} = \sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)}\boldsymbol{u}_t^{(i)\intercal}\right] & \tilde{U}_{(1,T)}\sum_{i=1}^{W}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)}x_t^{(i)\intercal}\right] & \bar{U}^{(i)} = \sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{u}_t^{(i)\intercal}\right]
\end{array} \tag{5}
$$

$$
L_y = -\frac{1}{2}\left[\begin{array}{c}
\mathrm{Tr}\left[R^{-1}Y\right] \\
-\mathrm{Tr}\left[R^{-1}C\tilde{Y}\right] - \mathrm{Tr}\left[R^{-1}\tilde{Y}^\intercal C^\intercal\right] \\
-\mathrm{Tr}\left[R^{-1}DU_y\right] - \mathrm{Tr}\left[R^{-1}U_y^\intercal D^\intercal\right] \\
-\mathrm{Tr}\left[R^{-1}\sum_{i=1}^W \boldsymbol{d}^{(i)}\bar{Y}^{(i)\intercal}\right] - \mathrm{Tr}\left[R^{-1}\sum_{i=1}^W \bar{Y}^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}CM_{(1,T)}C^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}D\tilde{U}_{(1,T)}C^\intercal\right] + \mathrm{Tr}\left[R^{-1}C\tilde{U}_{(1,T)}^\intercal D^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^W \boldsymbol{d}^{(i)}\bar{M}^{(i)\intercal}C^\intercal\right] + \mathrm{Tr}\left[R^{-1}C\sum_{i=1}^W \bar{M}^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}DU_{(1,T)}D^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^W \boldsymbol{d}^{(i)}\bar{U}^{(i)\intercal}D^\intercal\right] + \mathrm{Tr}\left[R^{-1}D\sum_{i=1}^W \bar{U}^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^W T^{(i)}\boldsymbol{d}^{(i)}\boldsymbol{d}^{(i)\intercal}\right]
\end{array}\right] - \frac{1}{2}\sum_{i=1}^W T^{(i)}\log|2\pi R| \tag{6}
$$

Finally, add a lagrange multiplier such that the columns of C sum to the 1 vector. $-\boldsymbol{\lambda}^\intercal(1 - C1) = -\boldsymbol{\lambda}^\intercal 1 + \boldsymbol{\lambda}^\intercal C1$

$$
L_y = -\frac{1}{2}\left[\begin{array}{c}
\mathrm{Tr}\left[R^{-1}Y\right] \\
-\mathrm{Tr}\left[R^{-1}C\tilde{Y}\right] - \mathrm{Tr}\left[R^{-1}\tilde{Y}^\intercal C^\intercal\right] \\
-\mathrm{Tr}\left[R^{-1}DU_y\right] - \mathrm{Tr}\left[R^{-1}U_y^\intercal D^\intercal\right] \\
-\mathrm{Tr}\left[R^{-1}\sum_{i=1}^W \boldsymbol{d}^{(i)}\bar{Y}^{(i)\intercal}\right] - \mathrm{Tr}\left[R^{-1}\sum_{i=1}^W \bar{Y}^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}CM_{(1,T)}C^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}D\tilde{U}_{(1,T)}C^\intercal\right] + \mathrm{Tr}\left[R^{-1}C\tilde{U}_{(1,T)}^\intercal D^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^W \boldsymbol{d}^{(i)}\bar{M}^{(i)\intercal}C^\intercal\right] + \mathrm{Tr}\left[R^{-1}C\sum_{i=1}^W \bar{M}^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}DU_{(1,T)}D^\intercal\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^W \boldsymbol{d}^{(i)}\bar{U}^{(i)\intercal}D^\intercal\right] + \mathrm{Tr}\left[R^{-1}D\sum_{i=1}^W \bar{U}^{(i)}\boldsymbol{d}^{(i)\intercal}\right] \\
+\mathrm{Tr}\left[R^{-1}\sum_{i=1}^W T^{(i)}\boldsymbol{d}^{(i)}\boldsymbol{d}^{(i)\intercal}\right]
\end{array}\right] - \frac{1}{2}\sum_{i=1}^W T^{(i)}\log|2\pi R| - \boldsymbol{\lambda}^\intercal 1 + \boldsymbol{\lambda}^\intercal C1 \tag{7}
$$

Take the derivative of L_y with respect to C, $\boldsymbol{\lambda}$, and $\boldsymbol{d}^{(i)}$ using the trace derivative equations in the appendix. Treat D as a constant. Note that all the transposed terms end up with the same derivative. Don't forget the $-\frac{1}{2}$ in front

$$
\begin{array}{rcc}
\frac{\partial L_y}{\partial C} & = & R^{-1}\tilde{Y}^\intercal - R^{-1}CM_{(1,T)} - R^{-1}D\tilde{U}_{(1,T)} - R^{-1}\sum_{i=1}^W \boldsymbol{d}^{(i)}\bar{M}^{(i)\intercal} + \boldsymbol{\lambda}1^\intercal \\
\frac{\partial L_y}{\partial \boldsymbol{\lambda}} & = & C1 - 1 \\
\frac{\partial L_y}{\partial \boldsymbol{d}^{(i)}} & = & R^{-1}\bar{Y}^{(i)} - R^{-1}C\bar{M}^{(i)} - R^{-1}D\bar{U}^{(i)} - T^{(i)}R^{-1}\boldsymbol{d}^{(i)}
\end{array} \tag{8}
$$

Set all terms to 0, multiply by R and gather terms that don't include C, λ, or $d^{(i)}$

$$
\begin{array}{rcc}
\tilde{Y}^\intercal - D\tilde{U}_{(1,T)} & = & CM_{(1,T)} + \sum_{i=1}^W \boldsymbol{d}^{(i)}\bar{M}^{(i)\intercal} - R\boldsymbol{\lambda}1^\intercal \\
1 & = & C1 \\
\bar{Y}^{(i)} - D\bar{U}^{(i)} & = & C\bar{M}^{(i)} + T^{(i)}\boldsymbol{d}^{(i)}
\end{array} \tag{9}
$$

Transpose and rewrite as a linear equation. Because R is diagonal, we could solve this equation column by column and treat R like a scalar. However, all it does it change the amplitude of $\boldsymbol{\lambda}$, without affecting any other parameter. We don't use $\boldsymbol{\lambda}$ anywhere so we can safely remove the $-R$ term.

$$
\begin{array}{rcc}
\tilde{Y} - \tilde{U}_{(1,T)}^\intercal D^\intercal & = & M_{(1,T)}C^\intercal + \bar{M}^{(i)}\sum_{i=1}^W \boldsymbol{d}^{(i)\intercal} + 1\boldsymbol{\lambda}^\intercal(-R) \\
1^\intercal & = & 1^\intercal C^\intercal \\
\bar{Y}^{(i)\intercal} - \bar{U}^{(i)\intercal}D^\intercal & = & \bar{M}^{(i)\intercal}C^\intercal + T^{(i)}\boldsymbol{d}^{(i)\intercal}
\end{array} \tag{10}
$$

$$
\begin{bmatrix} \tilde{Y} - \tilde{U}_{(1,T)}^{\intercal} D^{\intercal} \\ 1^{\intercal} \\ \bar{Y}^{(i)\intercal} - \bar{U}^{(i)\intercal} D^{\intercal} \end{bmatrix} = \begin{bmatrix} M_{(1,T)} & 1 & \bar{M}^{(1)} & \cdots & \bar{M}^{(W)} \\ 1^{\intercal} & 0 & 0 & 0 & 0 \\ \bar{M}^{(1)\intercal} & 0 & T^{(1)} & 0 & 0 \\ \vdots & 0 & 0 & \ddots & 0 \\ \bar{M}^{(W)\intercal} & 0 & 0 & 0 & T^{(W)} \end{bmatrix} \begin{bmatrix} C^{\intercal} \\ \boldsymbol{\lambda}^{\intercal} \\ \boldsymbol{d}^{(1)\intercal} \\ \vdots \\ \boldsymbol{d}^{(W)\intercal} \end{bmatrix} \tag{11}
$$

Here is an example with two data sets

$$
\begin{bmatrix} \tilde{Y} - \tilde{U}_{(1,T)}^{\intercal} D^{\intercal} \\ 1^{\intercal} \\ \bar{Y}^{(1)\intercal} - \bar{U}^{(1)\intercal} D^{\intercal} \\ \bar{Y}^{(2)\intercal} - \bar{U}^{(2)\intercal} D^{\intercal} \end{bmatrix} = \begin{bmatrix} M_{(1,T)} & 1 & \bar{M}^{(1)} & \bar{M}^{(2)} \\ 1^{\intercal} & 0 & 0 & 0 \\ \bar{M}^{(1)\intercal} & 0 & T^{(1)} & 0 \\ \bar{M}^{(2)\intercal} & 0 & 0 & T^{(2)} \end{bmatrix} \begin{bmatrix} C^{\intercal} \\ \boldsymbol{\lambda}^{\intercal} \\ \boldsymbol{d}^{(1)\intercal} \\ \boldsymbol{d}^{(2)\intercal} \end{bmatrix} \tag{12}
$$

Solving for R

Derivative rules are in the appendix

$$
\frac{\partial L_y}{\partial R^{-1}} = -\frac{1}{2} \begin{bmatrix} Y \\ -C\tilde{Y} - \tilde{Y}^{\intercal} C^{\intercal} \\ -DU_y - U_y^{\intercal} D^{\intercal} \\ -\sum_{i=1}^{W} \boldsymbol{d}^{(i)} \bar{Y}^{(i)\intercal} - \sum_{i=1}^{W} \bar{Y}^{(i)} \boldsymbol{d}^{(i)\intercal} \\ +CM_{(1,T)} C^{\intercal} \\ +D\tilde{U}_{(1,T)} C^{\intercal} + C\tilde{U}_{(1,T)}^{\intercal} D^{\intercal} \\ +\sum_{i=1}^{W} \boldsymbol{d}^{(i)} \bar{M}^{(i)\intercal} C^{\intercal} + C \sum_{i=1}^{W} \bar{M}^{(i)} \boldsymbol{d}^{(i)\intercal} \\ +DU_{(1,T)} D^{\intercal} \\ +\sum_{i=1}^{W} \boldsymbol{d}^{(i)} \bar{U}^{(i)\intercal} D^{\intercal} + D \sum_{i=1}^{W} \bar{U}^{(i)} \boldsymbol{d}^{(i)\intercal} \\ +\sum_{i=1}^{W} T^{(i)} \boldsymbol{d}^{(i)} \boldsymbol{d}^{(i)\intercal} \end{bmatrix} + \frac{1}{2} \sum_{i=1}^{W} T^{(i)} R \tag{13}
$$

Set the derivative to 0 and solve for R

$$
R = \frac{1}{\sum_{i=1}^{W} T^{(i)}} \begin{bmatrix} Y \\ -C\tilde{Y} - \tilde{Y}^{\intercal} C^{\intercal} \\ -DU_y - U_y^{\intercal} D^{\intercal} \\ -\sum_{i=1}^{W} \boldsymbol{d}^{(i)} \bar{Y}^{(i)\intercal} - \sum_{i=1}^{W} \bar{Y}^{(i)} \boldsymbol{d}^{(i)\intercal} \\ +CM_{(1,T)} C^{\intercal} \\ +D\tilde{U}_{(1,T)} C^{\intercal} + C\tilde{U}_{(1,T)}^{\intercal} D^{\intercal} \\ +\sum_{i=1}^{W} \boldsymbol{d}^{(i)} \bar{M}^{(i)\intercal} C^{\intercal} + C \sum_{i=1}^{W} \bar{M}^{(i)} \boldsymbol{d}^{(i)\intercal} \\ +DU_{(1,T)} D^{\intercal} \\ +\sum_{i=1}^{W} \boldsymbol{d}^{(i)} \bar{U}^{(i)\intercal} D^{\intercal} + D \sum_{i=1}^{W} \bar{U}^{(i)} \boldsymbol{d}^{(i)\intercal} \\ +\sum_{i=1}^{W} T^{(i)} \boldsymbol{d}^{(i)} \boldsymbol{d}^{(i)\intercal} \end{bmatrix} \tag{14}
$$

Appendix

Trace and determinant derivatives

$$
\frac{\partial}{\partial X} \mathrm{Tr}\left[AXB\right] = A^{\intercal} B^{\intercal} \tag{15}
$$

$$
\frac{\partial}{\partial X} \mathrm{Tr}\left[AX^{\intercal} B\right] = BA \tag{16}
$$

If C is the identity and A and B are symmetric

$$
\frac{\partial}{\partial X} \mathrm{Tr}\left[A_{\mathrm{sym}} X B_{\mathrm{sym}} X^{\intercal}\right] = 2AXB \tag{17}
$$

$$\frac{\partial}{\partial X^{-1}} \log |X| = -X \tag{18}$$