Supplementary material for "Analyzing Free-standing Conversational Groups A Multimodal Approach": Derivations of the ADMM

The minimization of the augmented Lagrangian with respect to $\mathbf{J}_\mathrm{b}$ and $\mathbf{J}_\mathrm{h}$, *i.e.* (9) can be split into two independent problems:

$$\mathbf{J}_\mathrm{b}^{r+1} = \operatorname*{argmin}_{\mathbf{J}_\mathrm{b}} \nu_\mathrm{b} \|\mathbf{J}_\mathrm{b}\|_* + \frac{\phi_\mathrm{b}}{2}\|\mathbf{K}_\mathrm{b}^{r+1} - \mathbf{J}_\mathrm{b}\|_\mathcal{F}^2 + \langle \mathbf{M}_\mathrm{b}^{r+1}, \mathbf{J}_\mathrm{b} - \mathbf{K}_\mathrm{b}^{r+1}\rangle$$

$$\mathbf{J}_\mathrm{h}^{r+1} = \operatorname*{argmin}_{\mathbf{J}_\mathrm{h}} \nu_\mathrm{h} \|\mathbf{J}_\mathrm{h}\|_* + \frac{\phi_\mathrm{h}}{2}\|\mathbf{K}_\mathrm{h}^{r+1} - \mathbf{J}_\mathrm{h}\|_\mathcal{F}^2 + \langle \mathbf{M}_\mathrm{h}^{r+1}, \mathbf{J}_\mathrm{h} - \mathbf{K}_\mathrm{h}^{r+1}\rangle$$

These two optimization problems are strictly equivalent to:

$$\mathbf{J}_\mathrm{b}^{r+1} = \operatorname*{argmin}_{\mathbf{J}_\mathrm{b}} \frac{\nu_\mathrm{b}}{\phi_\mathrm{b}} \|\mathbf{J}_\mathrm{b}\|_* + \frac{1}{2}\left\| \frac{1}{\phi_\mathrm{b}}\mathbf{M}_\mathrm{b}^{r+1} + \mathbf{J}_\mathrm{b} - \mathbf{K}_\mathrm{b}^{r+1}\right\|_\mathcal{F}^2,$$

$$\mathbf{J}_\mathrm{h}^{r+1} = \operatorname*{argmin}_{\mathbf{J}_\mathrm{h}} \frac{\nu_\mathrm{h}}{\phi_\mathrm{h}} \|\mathbf{J}_\mathrm{h}\|_* + \frac{1}{2}\left\| \frac{1}{\phi_\mathrm{h}}\mathbf{M}_\mathrm{h}^{r+1} + \mathbf{J}_\mathrm{h} - \mathbf{K}_\mathrm{h}^{r+1}\right\|_\mathcal{F}^2,$$

which are solved in closed-form by computing the singular value decomposition of $\mathbf{K}_\mathrm{b}^{r+1} - \frac{1}{\phi_\mathrm{b}}\mathbf{M}_\mathrm{b}^{r+1}$ and shrinking its singular values with the factor $\frac{\nu_\mathrm{b}}{\phi_\mathrm{b}}$ [7].

The optimization with respect to $\mathbf{K}_\mathrm{b}$ and $\mathbf{K}_\mathrm{h}$, *i.e.* (10) is slightly more complicated. The idea is to compute the gradient, and then find the critical point by solving a linear system of equations, so to cancel out the gradient. In order to do that, we first notice that the nuclear norms depend neither on $\mathbf{K}_\mathrm{b}$ nor on $\mathbf{K}_\mathrm{h}$. Furthermore, the derivative of the regularization terms added when writting the augmented Lagrangian, that is the last four terms of (8), is trivial. We focus in the derivatives of the error terms, the temporal regularization terms and the coupling terms, all of them including projections.

In order to to that, we rewrite these terms with the row-vectorized matrices, $\mathbf{k}_\mathrm{b} = \mathrm{vec}(\mathbf{K}_\mathrm{b})$ and $\mathbf{k}_\mathrm{h} = \mathrm{vec}(\mathbf{K}_\mathrm{h})$:

$$\left\|P_\mathcal{O}^\mathrm{b}\left(\widetilde{\mathbf{J}}_\mathrm{b} - \mathbf{K}_\mathrm{b}\right)\right\|_\mathcal{F}^2 = \left\|\mathbf{P}_\mathcal{O}^\mathrm{b}\left(\widetilde{\mathbf{j}}_\mathrm{b} - \mathbf{k}_\mathrm{b}\right)\right\|^2, \tag{19}$$

$$\mathrm{Tr}\left(P_\Theta^\mathrm{b}(\mathbf{K}_\mathrm{b})^\top \mathbf{T}_\mathrm{b} P_\Theta^\mathrm{b}(\mathbf{K}_\mathrm{b})\right) = \mathbf{k}_\mathrm{b}^\top \mathbf{P}_\Theta^{\mathrm{b}\top}\left(\mathbf{I}_C \otimes \mathbf{T}_\mathrm{b}\right)\mathbf{P}_\Theta^\mathrm{b}\mathbf{k}_\mathrm{b}, \tag{20}$$

$$\left\|P_\Theta^\mathrm{b}(\mathbf{K}_\mathrm{b}) - P_\Theta^\mathrm{h}(\mathbf{K}_\mathrm{h})\right\|_\mathcal{F}^2 = \left\|\mathbf{P}_\Theta^\mathrm{b}\mathbf{k}_\mathrm{b} - \mathbf{P}_\Theta^\mathrm{h}\mathbf{k}_\mathrm{h}\right\|^2, \tag{21}$$

where $\widetilde{\mathbf{j}}_\mathrm{b} = \mathrm{vec}(\widetilde{\mathbf{J}}_\mathrm{b})$, and $\mathbf{P}_\Theta^\mathrm{b} = [\mathbf{I}_{KTC}\ \mathbf{0}_{KTC\times KT(d_b+1)}]$, and

$$\mathbf{P}_\mathcal{O}^\mathrm{b} = \begin{bmatrix} \mathbf{I}_{KC}\otimes[\mathbf{I}_{T_0}\ \mathbf{0}_{T_0\times T-T_0}] & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{KT(d_b+1)} \end{bmatrix}. \tag{22}$$

With these notations the gradient of the objective function in (10), $\mathfrak{L}_{(10)} = \mathfrak{L}(\mathbf{J}_\mathrm{b}^{r+1}, \mathbf{J}_\mathrm{h}^{r+1}, \mathbf{K}_\mathrm{b}, \mathbf{K}_\mathrm{h}, \mathbf{M}_\mathrm{b}^r, \mathbf{M}_\mathrm{h}^r)$ with respect to $\mathbf{k}_\mathrm{b}$ and $\mathbf{k}_\mathrm{h}$ writes:

$$\begin{aligned}\frac{\partial \mathfrak{L}_{(10)}}{\partial \mathbf{k}_\mathrm{b}} &= \lambda_\mathrm{b}\mathbf{P}_\mathcal{O}^{\mathrm{b}\top}\mathbf{P}_\mathcal{O}^\mathrm{b}\left(\mathbf{k}_\mathrm{b} - \widetilde{\mathbf{j}}_\mathrm{b}\right) + \tau_\mathrm{b}\mathbf{P}_\Theta^{\mathrm{b}\top}\left(\mathbf{I}_C\otimes\mathbf{T}_\mathrm{b}\right)\mathbf{P}_\Theta^\mathrm{b}\mathbf{k}_\mathrm{b} \\ &+ \mathbf{P}_\Theta^{\mathrm{b}\top}\left(\mathbf{P}_\Theta^\mathrm{b}\mathbf{k}_\mathrm{b} - \mathbf{P}_\Theta^\mathrm{h}\mathbf{k}_\mathrm{h}\right) - \mathbf{m}_\mathrm{b}^r + \phi_\mathrm{b}\left(\mathbf{k}_\mathrm{b} - \mathbf{j}_\mathrm{b}^{r+1}\right),\end{aligned} \tag{23}$$

$$\begin{aligned}\frac{\partial \mathfrak{L}_{(10)}}{\partial \mathbf{k}_\mathrm{h}} &= \lambda_\mathrm{h}\mathbf{P}_\mathcal{O}^{\mathrm{h}\top}\mathbf{P}_\mathcal{O}^\mathrm{h}\left(\mathbf{k}_\mathrm{h} - \widetilde{\mathbf{j}}_\mathrm{h}\right) + \tau_\mathrm{h}\mathbf{P}_\Theta^{\mathrm{h}\top}\left(\mathbf{I}_C\otimes\mathbf{T}_\mathrm{h}\right)\mathbf{P}_\Theta^\mathrm{h}\mathbf{k}_\mathrm{h} \\ &+ \mathbf{P}_\Theta^{\mathrm{h}\top}\left(\mathbf{P}_\Theta^\mathrm{h}\mathbf{k}_\mathrm{h} - \mathbf{P}_\Theta^\mathrm{b}\mathbf{k}_\mathrm{b}\right) - \mathbf{m}_\mathrm{h}^r + \phi_\mathrm{h}\left(\mathbf{k}_\mathrm{h} - \mathbf{j}_\mathrm{h}^{r+1}\right),\end{aligned} \tag{24}$$

where the bold-faced lower-case letters stand for the row-vectorization of the respective matrices. Before proceeding we notice that the following relations hold:

$$\mathbf{P}_\mathcal{O}^{\mathrm{b}\top}\mathbf{P}_\mathcal{O}^\mathrm{b} = \begin{bmatrix} \mathbf{I}_{KC}\otimes\begin{bmatrix}\mathbf{I}_{T_0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}_{T-T_0}\end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{KT(d_b+1)}\end{bmatrix},$$

$$\mathbf{P}_\Theta^{\mathrm{b}\top}\left(\mathbf{I}_C\otimes\mathbf{T}_\mathrm{b}\right)\mathbf{P}_\Theta^\mathrm{b} = \begin{bmatrix}\mathbf{I}_{KC}\otimes\mathbf{L} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}_{KT(d_b+1)}\end{bmatrix},$$

$$\mathbf{P}_\Theta^{\mathrm{h}\top}\mathbf{P}_\Theta^\mathrm{b} = \begin{bmatrix}\mathbf{I}_{KTC} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \in \mathbb{R}^{KT(d_h+1+C)\times KT(d_b+1+C)}.$$

By canceling (23) and (24) we obtain:

$$\mathbf{Z}_\mathrm{b}\mathbf{k}_\mathrm{b} = \lambda_\mathrm{b}\mathbf{P}_\mathcal{O}^{\mathrm{b}\top}\mathbf{P}_\mathcal{O}^\mathrm{b}\widetilde{\mathbf{j}}_\mathrm{b} + \mathbf{P}_\Theta^{\mathrm{b}\top}\mathbf{P}_\Theta^\mathrm{h}\mathbf{k}_\mathrm{h} + \mathbf{m}_\mathrm{b}^r + \phi_\mathrm{b}\mathbf{j}_\mathrm{b}^{r+1} \tag{25}$$

$$\mathbf{Z}_\mathrm{h}\mathbf{k}_\mathrm{h} = \lambda_\mathrm{h}\mathbf{P}_\mathcal{O}^{\mathrm{h}\top}\mathbf{P}_\mathcal{O}^\mathrm{h}\widetilde{\mathbf{j}}_\mathrm{h} + \mathbf{P}_\Theta^{\mathrm{h}\top}\mathbf{P}_\Theta^\mathrm{b}\mathbf{k}_\mathrm{b} + \mathbf{m}_\mathrm{h}^r + \phi_\mathrm{h}\mathbf{j}_\mathrm{h}^{r+1} \tag{26}$$

with

$$\mathbf{Z}_\mathrm{b} = \begin{bmatrix}\mathbf{I}_{KC}\otimes\mathbf{L}_\mathrm{b} & \mathbf{0}\\ \mathbf{0} & (\lambda_\mathrm{b}+\phi_\mathrm{b})\mathbf{I}_{KT(d_b+1)}\end{bmatrix} \tag{27}$$

being $\mathbf{L}_\mathrm{b} = \lambda_\mathrm{b}\begin{bmatrix}\mathbf{I}_{T_0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}_{T-T_0}\end{bmatrix} + \tau_\mathrm{b}\mathbf{L} + (1+\phi_b)\mathbf{I}_T$ and analogously for $\mathbf{Z}_\mathrm{h}$.

Since the matrices of the systems above are block diagonal we can rewrite the system as several smaller systems. More precisely $\mathbf{Z}_\mathrm{b}$ (and $\mathbf{Z}_\mathrm{h}$) decompose in KC blocks of size T, plus a scaled identity matrix of size $KT(d_\mathrm{b}+1)$ (respectively $KT(d_\mathrm{h}+1)$ for the head). Importantly, the first part corresponds to $\mathbf{\Theta}^\mathrm{b}$ ($\mathbf{\Theta}^\mathrm{h}$) and the second part to $\mathbf{V}^\mathrm{b}$ ($\mathbf{V}^\mathrm{h}$).

We recall the notations $\mathbf{k}_{\mathrm{b},kc}$ and $\mathbf{k}_{\mathrm{b},\mathbf{v}}$ defined in Section 3.2. With these notations we can now easily rewrite the two previous systems of equations. The systems for $\mathbf{k}_{\mathrm{b},kc}$ and $\mathbf{k}_{\mathrm{h},kc}$ are coupled:

$$\mathbf{L}_\mathrm{b}\mathbf{k}_{\mathrm{b},kc} = \mathbf{k}_{\mathrm{h},kc} + \bar{\mathbf{k}}_{\mathrm{b},kc} \tag{28}$$

$$\mathbf{L}_\mathrm{h}\mathbf{k}_{\mathrm{h},kc} = \mathbf{k}_{\mathrm{b},kc} + \bar{\mathbf{k}}_{\mathrm{h},kc} \tag{29}$$

where $\bar{\mathbf{k}}_{\mathrm{b},kc}$ is defined as:

$$\bar{\mathbf{k}}_{\mathrm{b},kc} = \lambda_\mathrm{b}\begin{bmatrix}\mathbf{I}_{T_0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\tilde{\mathbf{j}}_{\mathrm{b},kc} + \mathbf{m}_{\mathrm{b},kc}^r + \phi_\mathrm{b}\mathbf{j}_{\mathrm{b},kc}^{r+1} \tag{30}$$

and $\bar{\mathbf{k}}_{\mathrm{h},kc}$ analogously for the head. The solution of this linear system is directly (14) and (15). The system for $\mathbf{k}_{\mathrm{b},\mathbf{v}}$ and $\mathbf{k}_{\mathrm{h},\mathbf{v}}$ is trivial and reduces directly to (16) and (17).