Head and Body Orientation Estimation with Sparse Weak Labels in Free Standing Conversational Settings

Supplementary Material

Stephanie Tan s.tan-1@tudelft.nl
David M.J. Tax d.m.j.tax@tudelft.nl
Hayley Hung h.hung@tudelft.nl
Delft University of Technology

Appendix A. Derivation of ADMM Update Rules

We start with the construction of the augmented Lagrangian of (1) (same as (10) in the manuscript and repeated here for convenience), similar to the classical ADMM (Eckstein and Yao, 2012).

$$\begin{aligned}
&\boldsymbol{J}_h^*, \boldsymbol{J}_b^* \\
&= \arg\min_{\boldsymbol{J}_h, \boldsymbol{J}_b} \underbrace{\nu_h \|\boldsymbol{J}_h\|_* + \nu_b \|\boldsymbol{J}_b\|_*}_{\text{matrix low-rankedness}} \\
&\quad + \underbrace{\frac{\lambda_h}{2}\|\boldsymbol{P}_h(\boldsymbol{J}_h - \boldsymbol{J}_{\mathrm{GP},h})\|_F^2 + \frac{\lambda_b}{2}\|\boldsymbol{P}_b(\boldsymbol{J}_b - \boldsymbol{J}_{\mathrm{GP},b})\|_F^2}_{\text{temporal smoothing}} \\
&\quad + \underbrace{\frac{\gamma_h}{2}\|\boldsymbol{P}_{w,h}(\boldsymbol{J}_h - \boldsymbol{J}_{w,h})\|_F^2 + \frac{\gamma_b}{2}\|\boldsymbol{P}_{w,b}(\boldsymbol{J}_b - \boldsymbol{J}_{w,b})\|_F^2}_{\text{weak label regularization}} \\
&\quad + \underbrace{\frac{\mu}{2}\|\boldsymbol{P}_h\boldsymbol{J}_h - \boldsymbol{P}_b\boldsymbol{J}_b\|_F^2}_{\text{head-body coupling}}
\end{aligned} \tag{1}$$

The augmented Lagrangian of the optimization problem is given by

$$\begin{aligned}
\mathcal{L} = &\ \nu_h \|\boldsymbol{J}_h\|_* + \nu_b \|\boldsymbol{J}_b\|_* \\
&+ \frac{\lambda_h}{2}\|\boldsymbol{P}_h(\boldsymbol{K}_h - \boldsymbol{J}_{\mathrm{GP},h})\|_F^2 + \frac{\lambda_b}{2}\|\boldsymbol{P}_b(\boldsymbol{K}_b - \boldsymbol{J}_{\mathrm{GP},b})\|_F^2 \\
&+ \frac{\gamma_h}{2}\|\boldsymbol{P}_{w,h}(\boldsymbol{K}_h - \boldsymbol{J}_{w,h})\|_F^2 + \frac{\gamma_b}{2}\|\boldsymbol{P}_{w,b}(\boldsymbol{K}_b - \boldsymbol{J}_{w,b})\|_F^2 \\
&+ \frac{\mu}{2}\|\boldsymbol{P}_h\boldsymbol{K}_h - \boldsymbol{P}_b\boldsymbol{K}_b\|_F^2 \\
&+ \frac{\phi_h}{2}\|\boldsymbol{K}_h - \boldsymbol{J}_h\|_F^2 + \frac{\phi_b}{2}\|\boldsymbol{K}_b - \boldsymbol{J}_b\|_F^2 \\
&+ \langle \boldsymbol{M}_h, \boldsymbol{J}_h - \boldsymbol{K}_h \rangle + \langle \boldsymbol{M}_b, \boldsymbol{J}_b - \boldsymbol{K}_b \rangle,
\end{aligned} \tag{2}$$

where $\boldsymbol{K}_h$ and $\boldsymbol{K}_b$ are auxiliary variables that allow the decoupling of the optimization of $\boldsymbol{J}_h$ and $\boldsymbol{J}_b$; and $\boldsymbol{M}_h$ and $\boldsymbol{M}_b$ are Lagrange multiplier matrices. The operator $\langle \cdot, \cdot \rangle$ denotes the inner product of two matrices. The update rules are similar to those of the ADMM with scaled dual variables (Boyd et al., 2011), and modified for our application. In this context, the update rules at the k-th iteration are given by

$$\begin{aligned}(\boldsymbol{J}_h^{k+1}, \boldsymbol{J}_b^{k+1}) = \arg\min_{\boldsymbol{J}_h^k, \boldsymbol{J}_b^k} & \nu_h \|\boldsymbol{J}_h^k\|_* + \nu_b \|\boldsymbol{J}_b^k\|_* \\ & + \frac{\phi_h}{2}\|\boldsymbol{K}_h^k - \boldsymbol{J}_h^k\|_F^2 + \frac{\phi_b}{2}\|\boldsymbol{K}_b^k - \boldsymbol{J}_b^k\|_F^2 \\ & + \langle \boldsymbol{M}_h^k, \boldsymbol{J}_h^k - \boldsymbol{K}_h^k \rangle + \langle \boldsymbol{M}_b^k, \boldsymbol{J}_b^k - \boldsymbol{K}_b^k \rangle \end{aligned} \tag{3}$$

$$\begin{aligned}&(\boldsymbol{K}_h^{k+1}, \boldsymbol{K}_b^{k+1}) \\ &= \arg\min_{\boldsymbol{K}_h^k, \boldsymbol{K}_b^k} \frac{\mu}{2}\|\boldsymbol{P}_h\boldsymbol{K}_h^k - \boldsymbol{P}_b\boldsymbol{K}_b^k\|_F^2 \\ &+ \frac{\lambda_h}{2}\|\boldsymbol{P}_h(\boldsymbol{K}_h^k - \boldsymbol{J}_{\mathrm{GP},h})\|_F^2 + \frac{\lambda_b}{2}\|\boldsymbol{P}_b(\boldsymbol{K}_b^k - \boldsymbol{J}_{\mathrm{GP},b})\|_F^2 \\ &+ \frac{\gamma_h}{2}\|\boldsymbol{P}_{w,h}(\boldsymbol{K}_h^k - \boldsymbol{J}_{w,h})\|_F^2 + \frac{\gamma_b}{2}\|\boldsymbol{P}_{w,b}(\boldsymbol{K}_b^k - \boldsymbol{J}_{w,b})\|_F^2 \\ &+ \frac{\phi_h}{2}\|\boldsymbol{K}_h^k - \boldsymbol{J}_h^{k+1}\|_F^2 + \frac{\phi_b}{2}\|\boldsymbol{K}_b^k - \boldsymbol{J}_b^{k+1}\|_F^2 \\ &+ \langle \boldsymbol{M}_h^k, \boldsymbol{J}_h^{k+1} - \boldsymbol{K}_h^k \rangle + \langle \boldsymbol{M}_b^k, \boldsymbol{J}_b^{k+1} - \boldsymbol{K}_b^k \rangle \end{aligned} \tag{4}$$

$$\boldsymbol{M}_h^{k+1} = \boldsymbol{M}_h^k + \phi_h(\boldsymbol{J}_h^{k+1} - \boldsymbol{K}_h^{k+1}) \tag{5}$$

$$\boldsymbol{M}_b^{k+1} = \boldsymbol{M}_b^k + \phi_b(\boldsymbol{J}_b^{k+1} - \boldsymbol{K}_b^{k+1}) \tag{6}$$

We sketch the approach on how to solve for $(\boldsymbol{J}_h^{k+1}, \boldsymbol{J}_b^{k+1})$ here for completeness. Separating and rearranging the head and body expressions for $(\boldsymbol{J}_h^{k+1}, \boldsymbol{J}_b^{k+1})$ which are the k_{th} iteration, we have

$$\begin{aligned}\boldsymbol{J}_h^{k+1} &= \arg\min_{J_h} \frac{\nu_h}{\phi_h}\|\boldsymbol{J}_h\|_* \\ &+ \frac{1}{2}\left\|\frac{1}{\phi_h}\boldsymbol{M}_h^k + \boldsymbol{J}_h - \boldsymbol{K}_h^k\right\|_F^2 - \frac{1}{2\phi_h}\langle \boldsymbol{M}_h^k, \boldsymbol{M}_h^k \rangle,\end{aligned} \tag{7}$$

and analogously for $\boldsymbol{J}_b^{k+1}$. According to derivations by Cai et al. (2010) and Alameda-Pineda et al. (2015), the solution of (7) arises from singular value decomposition of matrix $\boldsymbol{K}_h^k - \frac{1}{\phi_h}\boldsymbol{M}_h^k$, resulting in

$$\boldsymbol{J}_h^{k+1} = \boldsymbol{U}_h S_{\frac{\nu_h}{\phi_h}}(\boldsymbol{D}_h)\boldsymbol{V}_h^\top, \tag{8}$$

where $S_{\frac{\nu_h}{\phi_h}}$ is the shrinkage operator on the diagonal matrix. It follows analogously for $\boldsymbol{J}_b^{k+1}$.

Lastly we obtain the update rules for $\boldsymbol{K}_h^{k+1}$ and $\boldsymbol{K}_b^{k+1}$ by taking the derivative of (2) with respect to $\boldsymbol{k}_h$ and $\boldsymbol{k}_b$, which are the row-vectorization form of $\boldsymbol{K}_h$ and $\boldsymbol{K}_b$. The expressions are as follows:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{k}_h} = \lambda_h(\boldsymbol{k}_h - \boldsymbol{j}_{GP,h}) + \gamma_h(\boldsymbol{k}_h - \boldsymbol{j}_{w,h}) + \mu \boldsymbol{P}_h^\top(\boldsymbol{P}_h\boldsymbol{k}_h - \boldsymbol{P}_b\boldsymbol{k}_b) + \phi_h(\boldsymbol{k}_h - \boldsymbol{j}_h^{k+1}) - \boldsymbol{m}_h^k, \tag{9}$$

and

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{k}_b} = \lambda_b(\boldsymbol{k}_b - \boldsymbol{j}_{GP,b}) + \gamma_b(\boldsymbol{k}_b - \boldsymbol{j}_{w,b}) + \mu \boldsymbol{P}_b^\top(\boldsymbol{P}_b\boldsymbol{k}_b - \boldsymbol{P}_h\boldsymbol{k}_h) + \phi_b(\boldsymbol{k}_b - \boldsymbol{j}_b^{k+1}) - \boldsymbol{m}_b^k, \tag{10}$$

Setting (9) and (10) to zero gives

$$(\lambda_h + \gamma_h + \mu \boldsymbol{P}_h^\top \boldsymbol{P}_h + \phi_h)\boldsymbol{k}_h^{k+1} = \lambda_h \boldsymbol{j}_{GP,h} + \gamma_h \boldsymbol{j}_{w,h} + \mu \boldsymbol{P}_h^\top \boldsymbol{P}_b \boldsymbol{k}_b + \phi_h \boldsymbol{j}_h^{k+1} + \boldsymbol{m}_h^k \tag{11}$$

and

$$(\lambda_b + \gamma_b + \mu \boldsymbol{P}_b^\top \boldsymbol{P}_b + \phi_b)\boldsymbol{k}_b^{k+1} = \lambda_b \boldsymbol{j}_{GP,b} + \gamma_b \boldsymbol{j}_{w,b} + \mu \boldsymbol{P}_b^\top \boldsymbol{P}_h \boldsymbol{k}_h + \phi_b \boldsymbol{j}_b^{k+1} + \boldsymbol{m}_b^k. \tag{12}$$

We can solve these equations using methods based on LU decompositions or iterative methods. For the sake of brevity, we refer interested readers to derivations by Alameda-Pineda et al. (2015).

$\boldsymbol{K}_h^{k+1}$, $\boldsymbol{K}_b^{k+1}$, $\boldsymbol{J}_h^{k+1}$, $\boldsymbol{J}_b^{k+1}$, $\boldsymbol{M}_h^{k+1}$, and $\boldsymbol{M}_b^{k+1}$ are initialized as zero matrices and updated until convergence as $\|\boldsymbol{J}_h^{k+1} - \boldsymbol{K}_h^{k+1}\| \to 0$ and $\|\boldsymbol{J}_b^{k+1} - \boldsymbol{K}_b^{k+1}\| \to 0$.

References

Xavier Alameda-Pineda, Yan Yan, Elisa Ricci, Oswald Lanz, and Nicu Sebe. Analyzing free-standing conversational groups: A multimodal approach. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, pages 5–14, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3459-4. doi: 10.1145/2733373.2806238. URL http://doi.acm.org/10.1145/2733373.2806238.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

J. Cai, E. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. doi: 10.1137/080738970. URL https://doi.org/10.1137/080738970.

Jonathan Eckstein and W Yao. Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results. *RUTCOR Research Reports*, 32:3, 2012.