The detailed optimization of the classic BACF tracker

Fei Zhang

May 26, 2021

The overall objective function of BACF[1] is:

$$E(\mathbf{h}) = \frac{1}{2}\sum_{j=1}^{\mathrm{T}}\left\|\mathbf{y}(j) - \sum_{k=1}^{\mathrm{K}}\mathbf{h}_k^{\mathsf{T}}\mathbf{P}\mathbf{x}_k[\Delta\tau_j]\right\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{\mathrm{K}}\|\mathbf{h}_k\|_2^2 \tag{1}$$

where $\mathbf{P} \in \mathbb{R}^{\mathrm{D}\times\mathrm{T}}(\mathrm{T} \gg \mathrm{D})$, $\mathbf{x}_k \in \mathbb{R}^{\mathrm{T}}$, $\mathbf{y} \in \mathbb{R}^{\mathrm{T}}$, $\mathbf{h} \in \mathbb{R}^{\mathrm{D}}$.

$$\begin{aligned} E(\mathbf{h}) &= \frac{1}{2}\sum_{j=1}^{\mathrm{T}}\left\|\mathbf{y}(j) - \sum_{k=1}^{\mathrm{K}}\mathbf{X}_k(\mathbf{h}_k^{\mathsf{T}}\mathbf{P})^{\mathsf{T}}\right\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{\mathrm{K}}\|\mathbf{h}_k\|_2^2 \\ &= \frac{1}{2}\sum_{j=1}^{\mathrm{T}}\left\|\mathbf{y}(j) - \sum_{k=1}^{\mathrm{K}}\mathbf{X}_k\mathbf{P}^{\mathsf{T}}\mathbf{h}_k^{\mathsf{T}}\right\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{\mathrm{K}}\|\mathbf{h}_k\|_2^2 \end{aligned} \tag{2}$$

where $\mathbf{X}_k = \begin{bmatrix} \mathbf{x}_k[\Delta\tau_{\mathrm{T}}]^{\mathsf{T}} \\ \mathbf{x}_k[\Delta\tau_1]^{\mathsf{T}} \\ \vdots \\ \mathbf{x}_k[\Delta\tau_{\mathrm{T}-1}]^{\mathsf{T}} \end{bmatrix}$.

Defining the auxiliary variable $\mathbf{g}_k = \mathbf{P}^\mathsf{T}\mathbf{h}_k \in \mathbb{R}^\mathrm{T}$,

$$\begin{aligned}
\mathcal{F}(\sum_{k=1}^{K}\mathbf{X}_k\mathbf{P}^\mathsf{T}\mathbf{h}_k) &= \mathcal{F}(\sum_{k=1}^{K}\mathbf{x}_k * \mathbf{P}^\mathsf{T}\mathbf{h}_k) \\
&= \mathcal{F}(\sum_{k=1}^{K}\hat{\mathbf{x}}_k * \widehat{\mathbf{P}^\mathsf{T}\mathbf{h}_k}) \\
&= \sum_{k=1}^{K}\begin{bmatrix}\hat{\mathbf{x}}_{k1}\hat{\mathbf{g}}_{k1}\\ \hat{\mathbf{x}}_{k2}\hat{\mathbf{g}}_{k2}\\ \vdots \\ \hat{\mathbf{x}}_{k\mathrm{T}}\hat{\mathbf{g}}_{k\mathrm{T}}\end{bmatrix} = \begin{bmatrix}\sum_{k=1}^{K}\hat{\mathbf{x}}_{k1}\hat{\mathbf{g}}_{k1}\\ \sum_{k=1}^{K}\hat{\mathbf{x}}_{k2}\hat{\mathbf{g}}_{k2}\\ \vdots \\ \sum_{k=1}^{K}\hat{\mathbf{x}}_{k\mathrm{T}}\hat{\mathbf{g}}_{k\mathrm{T}}\end{bmatrix} \\
&= [diag(\hat{\mathbf{x}}_1), diag(\hat{\mathbf{x}}_2), \cdots, diag(\hat{\mathbf{x}}_\mathrm{K})]\begin{bmatrix}\hat{\mathbf{g}}_1\\ \hat{\mathbf{g}}_2\\ \vdots \\ \hat{\mathbf{g}}_\mathrm{K}\end{bmatrix} \\
&= \hat{\mathbf{X}}\hat{\mathbf{g}}
\end{aligned} \tag{3}$$

where $\mathbf{g}_k = [\mathbf{g}_{k1}, \mathbf{g}_{k2}, \cdots, \mathbf{g}_{k\mathrm{T}}]^\mathsf{T} \in \mathbb{R}^\mathrm{T}$.

$$\begin{aligned}
\hat{\mathbf{g}} &= \begin{bmatrix}\hat{\mathbf{g}}_1\\ \hat{\mathbf{g}}_2\\ \vdots \\ \hat{\mathbf{g}}_\mathrm{K}\end{bmatrix} \\
&= \begin{bmatrix}\sqrt{\mathrm{T}}\mathbf{F}\mathbf{P}^\mathsf{T}\mathbf{h}_1\\ \sqrt{\mathrm{T}}\mathbf{F}\mathbf{P}^\mathsf{T}\mathbf{h}_2\\ \vdots \\ \sqrt{\mathrm{T}}\mathbf{F}\mathbf{P}^\mathsf{T}\mathbf{h}_\mathrm{K}\end{bmatrix} \\
&= \sqrt{\mathrm{T}}\mathbf{F}\begin{bmatrix}\mathbf{P}^\mathsf{T} & & & \\ & \mathbf{P}^\mathsf{T} & & \\ & & \ddots & \\ & & & \mathbf{P}^\mathsf{T}\end{bmatrix}\begin{bmatrix}\mathbf{h}_1\\ \mathbf{h}_2\\ \vdots \\ \mathbf{h}_\mathrm{K}\end{bmatrix} \\
&= \sqrt{\mathrm{T}}\mathbf{F}(\mathbf{I}_\mathrm{K} \otimes \mathbf{P}^\mathsf{T})\mathbf{h} \\
&= \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}
\end{aligned} \tag{4}$$

Eq. 2 can be transformed into Fourier domain after introducing auxiliary variable:

$$\begin{aligned}
&E(\mathbf{h}, \hat{\mathbf{g}}) = \frac{1}{2}\left\|\mathbf{y} - \hat{\mathbf{X}}\hat{\mathbf{g}}\right\|_2^2 + \frac{\lambda}{2}\|\mathbf{h}\|_2^2 \\
&s.t. \quad \hat{\mathbf{g}} = \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}
\end{aligned} \tag{5}$$

The ALM form of Eq. 5 is expressed as follows,

$$
\begin{aligned}
E(\mathbf{h}, \hat{\mathbf{g}}, \hat{\boldsymbol{\zeta}}) = \frac{1}{2\mathrm{T}} \left\| \mathbf{y} - \hat{\mathbf{X}}\hat{\mathbf{g}} \right\|_2^2 + \frac{\lambda}{2} \|\mathbf{h}\|_2^2 + \hat{\boldsymbol{\zeta}}^\mathsf{T}(\hat{\mathbf{g}} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}) \\
+ \frac{\mu}{2} \left\| \mathbf{g} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h} \right\|_2^2
\end{aligned} \tag{6}
$$

where $\hat{\boldsymbol{\zeta}} = [\boldsymbol{\zeta}_1^\mathsf{T}, \boldsymbol{\zeta}_2^\mathsf{T} \cdots, \boldsymbol{\zeta}_\mathrm{T}^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{\mathrm{KT}}$ and μ are the Lagrangian vector in the Fourier domain and the penalty factor, respectivly.

Then ADMM [2] algorithm is applied to Eq. 6,

$$
\begin{cases}
\mathbf{h}^* = \arg\min\limits_{\mathbf{h}} \frac{\lambda}{2} \|\mathbf{h}\|_2^2 + \hat{\boldsymbol{\zeta}}^\mathsf{T}(\hat{\mathbf{g}} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}) + \frac{\mu}{2} \left\| \mathbf{g} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h} \right\|_2^2 \\
\hat{\mathbf{g}}^* = \arg\min\limits_{\hat{\mathbf{g}}} \frac{1}{2\mathrm{T}} \left\| \hat{\mathbf{y}} - \hat{\mathbf{X}}\hat{\mathbf{g}} \right\|_2^2 + \hat{\boldsymbol{\zeta}}^\mathsf{T}(\hat{\mathbf{g}} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}) \\
\quad + \frac{\mu}{2} \left\| \mathbf{g} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h} \right\|_2^2 \\
\hat{\boldsymbol{\zeta}}^{(i+1)} = \hat{\boldsymbol{\zeta}}^{(i)} + \mu(\hat{\mathbf{g}}_t^{(i)} - \hat{\mathbf{h}}_t^{(i)}).
\end{cases} \tag{7}
$$

Subproblem h:

$$
\begin{aligned}
\mathbf{H} &= \frac{\lambda}{2}\mathbf{h}^\mathsf{T}\mathbf{h} + \hat{\boldsymbol{\zeta}}^\mathsf{T}\hat{\mathbf{g}} - \hat{\boldsymbol{\zeta}}^\mathsf{T}\sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h} \\
&+ \frac{\mu}{2}\left(\hat{\mathbf{g}} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}\right)^\mathsf{T}\left(\hat{\mathbf{g}} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}\right) \\
&= \frac{\lambda}{2}\mathbf{h}^\mathsf{T}\mathbf{h} + \hat{\boldsymbol{\zeta}}^\mathsf{T}\hat{\mathbf{g}} - \hat{\boldsymbol{\zeta}}^\mathsf{T}\sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h} \\
&+ \frac{\mu}{2}\left(\hat{\mathbf{g}} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}\right)^\mathsf{T}\left(\hat{\mathbf{g}} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h}\right) \\
&= \frac{\lambda}{2}\mathbf{h}^\mathsf{T}\mathbf{h} + \hat{\boldsymbol{\zeta}}^\mathsf{T}\hat{\mathbf{g}} - \hat{\boldsymbol{\zeta}}^\mathsf{T}\sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h} \\
&+ \frac{\mu}{2}\left(\hat{\mathbf{g}}^\mathsf{T}\hat{\mathbf{g}} - \sqrt{\mathrm{T}}\hat{\mathbf{g}}^\mathsf{T}(\mathbf{I}_\mathrm{K} \otimes \mathbf{F}\mathbf{P}^\mathsf{T})\mathbf{h} - \sqrt{\mathrm{T}}\mathbf{h}^\mathsf{T}(\mathbf{I}_\mathrm{K} \otimes \mathbf{P}\mathbf{F}^\mathsf{T})\hat{\mathbf{g}} + \mathrm{T}\mathbf{h}^\mathsf{T}\mathbf{h}\right)
\end{aligned} \tag{8}
$$

By differentiating the first subproblem in Eq. 7 with respect to the filter $\mathbf{h}$ and setting the outcome to zero, we can obtain,

$$
\lambda\mathbf{h} - \sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{P}\mathbf{F}^\mathsf{T})\hat{\boldsymbol{\zeta}} - \mu\sqrt{\mathrm{T}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{P}\mathbf{F}^\mathsf{T})\mathbf{h} + \mu\mathrm{T}\mathbf{h} = 0 \tag{9}
$$

Defining $g = \frac{1}{\sqrt{\mathrm{T}}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{P}\mathbf{F}^\mathsf{T})\hat{\mathbf{g}}$, $\boldsymbol{\zeta} = \frac{1}{\sqrt{\mathrm{T}}}(\mathbf{I}_\mathrm{K} \otimes \mathbf{P}\mathbf{F}^\mathsf{T})\hat{\boldsymbol{\zeta}}$,

$$
(\lambda + \mu\mathrm{T})\mathbf{h} = \mathrm{T}\boldsymbol{\zeta} + \mu\mathrm{T}\mathbf{g} \tag{10}
$$

Eq.10 can be simplified as,

$$
\mathbf{h}^* = \frac{\boldsymbol{\zeta} + \mu\mathbf{g}}{\lambda/\mathrm{T} + \mu} \tag{11}
$$

Subproblem $\hat{\mathbf{g}}$: The second subproblem in Eq. 7 can be decomposed into T smaller problems,

$$
\hat{\mathbf{g}}(t)^* = \frac{1}{2\mathrm{T}} \left\| \hat{\mathbf{y}}(t) - \hat{\mathbf{x}}(t)^\mathsf{T}\hat{\mathbf{g}}(t) \right\|_2^2 + \hat{\boldsymbol{\zeta}}^\mathsf{T}(\hat{\mathbf{g}}(t) - \hat{\mathbf{h}}(t)) + \frac{\mu}{2} \left\| \hat{\mathbf{g}}(t) - \hat{\mathbf{h}}(t) \right\|_2^2 \tag{12}
$$

$$
\begin{aligned}
\mathbf{G} &= \frac{1}{2\mathrm{T}}\left(\hat{\mathbf{y}}(t)-\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{g}}(t)\right)^{\mathsf{T}}\left(\hat{\mathbf{y}}(t)-\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{g}}(t)\right)+\hat{\boldsymbol{\zeta}}^{\mathsf{T}}\hat{\mathbf{g}}(t)-\hat{\boldsymbol{\zeta}}^{\mathsf{T}}\hat{\mathbf{h}}(t) \\
&+\frac{\mu}{2}\left(\hat{\mathbf{g}}(t)-\hat{\mathbf{h}}(t)\right)^{\mathsf{T}}\left(\hat{\mathbf{g}}(t)-\hat{\mathbf{h}}(t)\right) \\
&= \frac{1}{2\mathrm{T}}\left(\hat{\mathbf{y}}(t)^{\mathsf{T}}-\hat{\mathbf{g}}(t)^{\mathsf{T}}\hat{\mathbf{x}}(t)\right)\left(\hat{\mathbf{y}}(t)-\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{g}}(t)\right)+\hat{\boldsymbol{\zeta}}^{\mathsf{T}}\hat{\mathbf{g}}(t)-\hat{\boldsymbol{\zeta}}^{\mathsf{T}}\hat{\mathbf{h}}(t) \\
&+\frac{\mu}{2}\left(\hat{\mathbf{g}}(t)^{\mathsf{T}}-\hat{\mathbf{h}}(t)^{\mathsf{T}}\right)\left(\hat{\mathbf{g}}(t)-\hat{\mathbf{h}}(t)\right) \\
&= \frac{1}{2\mathrm{T}}\left(\hat{\mathbf{y}}(t)^{\mathsf{T}}\hat{\mathbf{y}}(t)-\hat{\mathbf{y}}(t)^{\mathsf{T}}\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{g}}(t)-\hat{\mathbf{g}}(t)^{\mathsf{T}}\hat{\mathbf{x}}(t)\hat{\mathbf{y}}(t)+\hat{\mathbf{g}}(t)^{\mathsf{T}}\hat{\mathbf{x}}(t)\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{g}}(t)\right) \\
&+\hat{\boldsymbol{\zeta}}^{\mathsf{T}}\hat{\mathbf{g}}(t)-\hat{\boldsymbol{\zeta}}^{\mathsf{T}}\hat{\mathbf{h}}(t)+\frac{\mu}{2}\left(\hat{\mathbf{g}}(t)^{\mathsf{T}}\hat{\mathbf{g}}(t)-\hat{\mathbf{g}}(t)^{\mathsf{T}}\hat{\mathbf{h}}(t)-\hat{\mathbf{h}}(t)^{\mathsf{T}}\hat{\mathbf{g}}(t)+\hat{\mathbf{h}}(t)^{\mathsf{T}}\hat{\mathbf{h}}(t)\right)
\end{aligned}
\tag{13}
$$

By differentiating Eq. 13 with respect to the filter $\hat{\mathbf{g}}$ and setting the outcome to zero, we can obtain,

$$
\frac{1}{2\mathrm{T}}\left(-2\hat{\mathbf{x}}(t)\hat{\mathbf{y}}(t)+2\hat{\mathbf{x}}(t)\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{g}}(t)\right)+\hat{\boldsymbol{\zeta}}(t)+\frac{\mu}{2}\left(2\hat{\mathbf{g}}(t)-2\hat{\mathbf{h}}(t)\right)=0 \tag{14}
$$

$$
\hat{\mathbf{g}}(t)^{*}=\left(\hat{\mathbf{x}}(t)\hat{\mathbf{x}}(t)^{\mathsf{T}}+\mu\mathrm{T}\mathbf{I}_{\mathrm{K}}\right)^{-1}\left(\hat{\mathbf{x}}(t)\hat{\mathbf{y}}(t)+\mu\mathrm{T}\hat{\mathbf{h}}(\mathrm{t})-\mathrm{T}\hat{\boldsymbol{\zeta}}(\mathrm{t})\right) \tag{15}
$$

Then, Sherman-Morrision[3] $(uv^{\mathrm{T}}+\mathrm{A})^{-1}=\mathrm{A}^{-1}-\frac{\mathrm{A}^{-1}uv^{\mathrm{T}}\mathrm{A}^{-1}}{1+v^{\mathrm{T}}\mathrm{A}^{-1}u}$ is used for acceleration, we get that,

$$
\left(\hat{\mathbf{x}}(t)\hat{\mathbf{x}}(t)^{\mathsf{T}}+\mu\mathrm{T}\mathbf{I}_{\mathrm{K}}\right)^{-1}=\frac{1}{\mu\mathrm{T}}\left(\mathrm{I}_{\mathrm{K}}-\frac{\hat{\mathbf{x}}(\mathrm{t})\hat{\mathbf{x}}(\mathrm{t})^{\mathsf{T}}}{\mu\mathrm{T}+\hat{\mathbf{x}}(\mathrm{t})^{\mathsf{T}}}\hat{\mathbf{x}}(\mathrm{t})\right) \tag{16}
$$

Final, Eq. 15 can be converted into,

$$
\begin{aligned}
\hat{\mathbf{g}}(t)^{*} &= \frac{1}{\mu\mathrm{T}}\left(\mathbf{I}_{\mathrm{K}}-\frac{\hat{\mathbf{x}}(t)\hat{\mathbf{x}}(t)^{\mathsf{T}}}{\mu\mathrm{T}}+\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{x}}(t)\right)\left(\hat{\mathbf{x}}(t)\hat{\mathbf{y}}(t)+\mu\mathrm{T}\hat{\mathbf{h}}(\mathrm{t})-\mathrm{T}\hat{\boldsymbol{\zeta}}(\mathrm{t})\right) \\
&= \frac{1}{\mu\mathrm{T}}\left(\hat{\mathbf{x}}(t)\hat{\mathbf{y}}(t)+\mu\mathrm{T}\hat{\mathbf{h}}(\mathrm{t})-\mathrm{T}\hat{\boldsymbol{\zeta}}(\mathrm{t})-\frac{\hat{\mathbf{x}}(\mathrm{t})\hat{\mathbf{x}}(\mathrm{t})^{\mathsf{T}}\hat{\mathbf{x}}(\mathrm{t})\hat{\mathbf{y}}(\mathrm{t})+\mu\mathrm{T}+\hat{\mathbf{x}}(\mathrm{t})\hat{\mathbf{x}}(\mathrm{t})^{\mathsf{T}}\hat{\mathbf{h}}(\mathrm{t})-\mathrm{T}\hat{\mathbf{x}}(\mathrm{t})\hat{\mathbf{x}}(\mathrm{t})^{\mathsf{T}}\hat{\boldsymbol{\zeta}}(\mathrm{t})}{\mu\mathrm{T}+\mathbf{x}(\mathrm{t})^{\mathsf{T}}\hat{\mathbf{x}}(\mathrm{t})}\right) \\
&= \frac{1}{\mu}\left(\frac{1}{\mathrm{T}}\hat{\mathbf{x}}(t)\hat{\mathbf{y}}(t)+\mu\hat{\mathbf{h}}(t)-\hat{\boldsymbol{\zeta}}(t)\right) \\
&- \frac{\hat{\mathbf{x}}(t)}{\mu b}\left(\frac{1}{\mathrm{T}}\hat{\mathbf{S}}_{x}(t)\hat{\mathbf{y}}(t)+\mu\hat{\mathbf{S}}_{h}(t)-\hat{\mathbf{S}}_{\boldsymbol{\zeta}}(t)\right)
\end{aligned}
\tag{17}
$$

where $\hat{\mathbf{S}}_{\mathbf{x}}(t)=\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{x}}(t)$,$\hat{\mathbf{S}}_{\mathbf{h}}(t)=\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\mathbf{x}}(t)$,$\hat{\mathbf{S}}_{\boldsymbol{\zeta}}(t)=\hat{\mathbf{x}}(t)^{\mathsf{T}}\hat{\boldsymbol{\zeta}}(t)$ and $b=\hat{\mathbf{S}}_{\mathbf{x}}(t)+\mu\mathrm{T}$.

Subproblem $\hat{\boldsymbol{\zeta}}$:

$$
\mu^{(i+1)}=min(\mu_{max},\beta\mu^{(i)}) \tag{18}
$$

A Equation

$$
\frac{d\mathbf{X}^{\mathsf{T}}}{d\mathbf{X}}=\mathbf{I} \tag{19}
$$

$$\frac{d\mathbf{X}}{d\mathbf{X}^\top} = \mathbf{I} \tag{20}$$

$$\frac{d\mathbf{A}\mathbf{X}}{d\mathbf{X}^\top} = \mathbf{A} \tag{21}$$

$$\frac{d\mathbf{X}^\top\mathbf{A}}{d\mathbf{X}} = \mathbf{A} \tag{22}$$

$$\frac{d\mathbf{A}\mathbf{X}}{d\mathbf{X}} = \mathbf{A}^\top \tag{23}$$

$$\frac{d\mathbf{X}\mathbf{A}}{d\mathbf{X}} = \mathbf{A}^\top \tag{24}$$

References

[1] H. Kiani Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1135–1143.

[2] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers.* Found. Trends Mach. Learn., 2011.

[3] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Stat.*, vol. 21, no. 1, pp. 124–127, 1950.