A Proofs

In this section, we prove Prop. 1. We start by stating and proving a lemma. Then, by rewriting Eq. 8, with $\mathcal{R} = \mathcal{R}_{\ell_2}$, in such a way that the stated lemma is applicable, we prove item 3 of Prop. 1. We then proceed to show that items 2 and 1 of Prop. 1 also hold. Finally, we prove that the analogous of Prop. 1 does not hold, when replacing the Euclidean norm by the ℓ_1 matrix norm.

A.1 Euclidean Norm

We start by proving the following lemma, which will be used to prove Prop. 1.

Lemma 3. *Assume $K \geq L$. Let matrices $\boldsymbol{M} \in \mathbb{R}^{V_{\mathrm{S}} \times V_{\mathrm{S}}}$ (invertible), $\boldsymbol{V} \in \mathbb{R}^{K \times L}$ (with full column rank) and $\boldsymbol{W} \in \mathbb{R}^{V_{\mathrm{S}} \times L}$ be arbitrary. Then, the matrix*

$$\boldsymbol{P}^* = \arg \min_{\boldsymbol{P}: \boldsymbol{P}\boldsymbol{V} = \boldsymbol{W}} \frac{1}{2} \|\boldsymbol{M}^\top \boldsymbol{P}\|_{\mathrm{F}}^2, \tag{11}$$

has rank at most L. Moreover, $\boldsymbol{P}^* = \boldsymbol{W}\boldsymbol{V}^\top (\boldsymbol{V}\boldsymbol{V}^\top)^{-1}$, regardless of $\boldsymbol{M}$.

Proof. Let $\otimes$ denote the Kronecker product and $\mathrm{vec}(.)$ the function that stacks together the columns of a matrix into a column vector. We use the well-known property (Petersen and Pedersen, 2012): for any matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{X}$ such that $\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$ is a valid product, the following holds:

$$(\boldsymbol{B}^\top \otimes \boldsymbol{A}) \,\mathrm{vec}(\boldsymbol{X}) = \mathrm{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}). \tag{12}$$

Then, we have the following:

$$\begin{aligned} \|\boldsymbol{M}^\top \boldsymbol{P}\|_{\mathrm{F}} &= \|\boldsymbol{P}^\top \boldsymbol{M}\|_{\mathrm{F}} = \|\mathrm{vec}(\boldsymbol{P}^\top \boldsymbol{M})\| \\ &= \|(\boldsymbol{M}^\top \otimes \boldsymbol{I}_K) \,\mathrm{vec}(\boldsymbol{P}^\top)\|. \end{aligned} \tag{13}$$

Furthermore, if we let $\boldsymbol{p} := \mathrm{vec}(\boldsymbol{P}^\top)$ and $\boldsymbol{w} := \mathrm{vec}(\boldsymbol{W}^\top)$, by the same reasoning, the following also holds:

$$\begin{aligned} &\mathrm{vec}(\boldsymbol{V}^\top \boldsymbol{P}^\top) = \mathrm{vec}(\boldsymbol{W}^\top) \\ \Leftrightarrow\ &(\boldsymbol{I}_{V_{\mathrm{S}}} \otimes \boldsymbol{V}^\top)\boldsymbol{p} = \boldsymbol{w}. \end{aligned} \tag{14}$$

Let $\boldsymbol{H} = (\boldsymbol{M}^\top \otimes \boldsymbol{I}_K)^\top (\boldsymbol{M}^\top \otimes \boldsymbol{I}_K) = (\boldsymbol{M} \otimes \boldsymbol{I}_K)(\boldsymbol{M}^\top \otimes \boldsymbol{I}_K) = \boldsymbol{M}\boldsymbol{M}^\top \otimes \boldsymbol{I}_K$ and $\boldsymbol{G} = \boldsymbol{I}_{V_{\mathrm{S}}} \otimes \boldsymbol{V}^\top$. Given the above properties, we can then pose the following optimization problem, which is equivalent to the minimization in Eq. 11:

$$\begin{aligned} \min_{\boldsymbol{p}} \quad & \frac{1}{2} \boldsymbol{p}^\top \boldsymbol{H} \boldsymbol{p} \\ \text{s.t.} \quad & \boldsymbol{G}\boldsymbol{p} = \boldsymbol{w}. \end{aligned} \tag{15}$$

This problem is equivalent to the following linear system (given by the Lagrangian conditions)

$$\begin{pmatrix} \boldsymbol{H} & \boldsymbol{G}^\top \\ \boldsymbol{G} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{w} \end{pmatrix}, \tag{16}$$

where λ is a vector of Lagrange multipliers. The solution to this system is

$$\begin{pmatrix} \boldsymbol{p}^* \\ \boldsymbol{\lambda}^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{H} & \boldsymbol{G}^\top \\ \boldsymbol{G} & \boldsymbol{0} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{w} \end{pmatrix}, \tag{17}$$

which gives

$$\begin{aligned} \boldsymbol{p}^* &= \boldsymbol{H}^{-1}\boldsymbol{G}^\top (\boldsymbol{G}\boldsymbol{H}^{-1}\boldsymbol{G}^\top)^{-1} \boldsymbol{w} \\ &= [\boldsymbol{I}_{V_{\mathrm{S}}} \otimes (\boldsymbol{V}(\boldsymbol{V}^\top \boldsymbol{V})^{-1})]\boldsymbol{w}, \end{aligned} \tag{18}$$

where the second line comes from expanding $\boldsymbol{G}$ and $\boldsymbol{H}$ and canceling several terms. Let $\boldsymbol{V}^+ = (\boldsymbol{V}^\top \boldsymbol{V})^{-1} \boldsymbol{V}^\top$ be the pseudo-inverse of $\boldsymbol{V}$. Again from the property of the Kronecker product in Eq. 12, we have that

$$\boldsymbol{P}^* = (\boldsymbol{V}(\boldsymbol{V}^\top \boldsymbol{V})^{-1} \boldsymbol{W}^\top)^\top = \boldsymbol{W}\boldsymbol{V}^+. \tag{19}$$

Note that this shows that the optimal $\boldsymbol{P}^*$ does not depend on $\boldsymbol{M}$. Since $\mathrm{rank}(\boldsymbol{V}^+) \leq L$, we have that $\mathrm{rank}(\boldsymbol{P}) \leq L$. Note that, in order to conclude this, we have not assumed anything about $\boldsymbol{M}$ or $\boldsymbol{V}$, other than that they are full row and column rank matrices, respectively. ☐

We now prove Prop. 1. Let $\boldsymbol{S}$ be the matrix whose columns are $\boldsymbol{s}^{(1)}, \ldots, \boldsymbol{s}^{(N)}$. If we keep $\boldsymbol{P}$ fixed and optimize only with respect to $\boldsymbol{Q}$, we obtain

$$\boldsymbol{Q}^* = \arg \min_{\boldsymbol{Q}} \frac{\mu}{2} \|\boldsymbol{S}^\top \boldsymbol{P} - \boldsymbol{T}^\top \boldsymbol{Q}\|_{\mathrm{F}}^2 + \frac{\mu_{\mathrm{T}}}{2} \|\boldsymbol{Q}\|_{\mathrm{F}}^2. \tag{20}$$

Setting the gradient to zero, and noting that $\boldsymbol{T}$ has full row rank, we obtain the following closed-form solution for $\boldsymbol{Q}^*$:

$$\boldsymbol{Q}^* = \left(\boldsymbol{T}\boldsymbol{T}^\top + \frac{\mu_{\mathrm{T}}}{\mu} \boldsymbol{I}_{V_{\mathrm{T}}} \right)^{-1} \boldsymbol{T}\boldsymbol{S}^\top \boldsymbol{P}. \tag{21}$$

The equation above can be written in the form $\boldsymbol{Q}^* = \boldsymbol{R}\boldsymbol{P}$, where $\boldsymbol{R} \in \mathbb{R}^{V_{\mathrm{T}} \times V_{\mathrm{S}}}$ (i.e., $\boldsymbol{Q}^*$ depends linearly on $\boldsymbol{P}$). Therefore, we can write the objective function in Eq. 8 (with $\mathcal{R} = \mathcal{R}_{\ell_2}$) as

$$\begin{aligned} \mathcal{F}(\boldsymbol{P}, \boldsymbol{Q}^*) = &\frac{\mu}{2} \|(\boldsymbol{S}^\top - \boldsymbol{T}^\top \boldsymbol{R})\boldsymbol{P}\|_{\mathrm{F}}^2 + \frac{\mu_{\mathrm{S}}}{2} \|\boldsymbol{P}\|_{\mathrm{F}}^2 \\ &+ \frac{\mu_{\mathrm{T}}}{2} \|\boldsymbol{R}\boldsymbol{P}\|_{\mathrm{F}}^2 + \mathcal{L}(\boldsymbol{P}\boldsymbol{V}). \end{aligned} \tag{22}$$

Note that the first three terms of Eq. 22 are all squared Frobenius norms of linear transformations of $\boldsymbol{P}$, hence we can collapse them all into a single term $\|\boldsymbol{M}^\top \boldsymbol{P}\|_\mathrm{F}^2$ for some matrix $\boldsymbol{M} \in \mathbb{R}^{V_\mathrm{S} \times V_\mathrm{S}}$. Finally, we rewrite our objective function as

$$\begin{aligned}&\min_{\boldsymbol{P}} \left(\|\boldsymbol{M}^\top \boldsymbol{P}\|_\mathrm{F}^2 + \mathcal{L}(\boldsymbol{PV})\right)\\ &= \min_{\boldsymbol{W}} \left[\left(\min_{\boldsymbol{P}:\boldsymbol{PV}=\boldsymbol{W}} \|\boldsymbol{M}^\top \boldsymbol{P}\|_\mathrm{F}^2\right) + \mathcal{L}(\boldsymbol{W})\right].\end{aligned} \tag{23}$$

Invoking Lemma 3 and the fact that $\boldsymbol{Q}^*$ is a linear transformation of $\boldsymbol{P}^*$, we have item 3 of Prop. 1. To prove item 2, we start by replacing Eq. 19 in Eq. 23, obtaining

$$\min_{\boldsymbol{W}} \|\boldsymbol{M}^\top \boldsymbol{W}\boldsymbol{V}^+\|_\mathrm{F}^2 + \mathcal{L}(\boldsymbol{W}). \tag{24}$$

We can simplify the quadratic term

$$\begin{aligned}\|\boldsymbol{M}^\top \boldsymbol{W}\boldsymbol{V}^+\|_\mathrm{F}^2 &= \|(\boldsymbol{M}^\top \otimes (\boldsymbol{V}^+)^\top)\boldsymbol{w}\|^2\\ &= \boldsymbol{w}^\top(\boldsymbol{M}\boldsymbol{M}^\top \otimes \boldsymbol{V}^+(\boldsymbol{V}^+)^\top)\boldsymbol{w}\\ &= \boldsymbol{w}^\top(\boldsymbol{M}\boldsymbol{M}^\top \otimes (\boldsymbol{V}^\top\boldsymbol{V})^{-1})\boldsymbol{w}.\end{aligned} \tag{25}$$

We can see from Eq. 25 that the classifier obtained by optimizing Eq. 24 depends on $\boldsymbol{V}$ only through the matrix product $\boldsymbol{V}^\top\boldsymbol{V}$, as stated in item 2 of Prop. 1.

We still need to show item 1 of Prop. 1. For any $\boldsymbol{V} \in \mathbb{R}^{K \times L}$ (that is full column rank), let $\boldsymbol{V}' \in \mathbb{R}^{K' \times L}$ be such that $K' = L$ and $\boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{V}'^\top\boldsymbol{V}'$, and let $\boldsymbol{W}^* = \boldsymbol{PV}$ be the minimizer of Eq. 24 and $\boldsymbol{W}'^* = \boldsymbol{P}'\boldsymbol{V}'$ the minimizer of the same expression, when using $\boldsymbol{V}'$ instead of $\boldsymbol{V}$. Since $\boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{V}'^\top\boldsymbol{V}'$, we have that $\boldsymbol{W}^* = \boldsymbol{W}'^*$. Then, by our definitions of $\boldsymbol{W}$ and $\boldsymbol{W}'$, we get

$$\boldsymbol{PV} = \boldsymbol{P}'\boldsymbol{V}'. \tag{26}$$

Hence, the classifier for the source language is the same when using $\boldsymbol{V}$ or $\boldsymbol{V}'$. A similar reasoning can be used to prove that $\boldsymbol{QV} = \boldsymbol{Q}'\boldsymbol{V}'$, and conclude that the classifier for the target language is also the same. This proves item 1 in Prop. 1, finishing our proof.

A.2 Generalization to Mahalanobis Norms

We define the Mahalanobis-Frobenius norm of a matrix $\boldsymbol{X} \in \mathbb{R}^{I \times J}$ induced by a positive definite matrix $\boldsymbol{R} \in \mathbb{R}^{I \times I}$ as $\|\boldsymbol{X}\|_{\boldsymbol{R}} := \sqrt{\sum_{j=1}^{J} \boldsymbol{x}_j^\top \boldsymbol{R}\boldsymbol{x}_j}$, where $\boldsymbol{x}_j$ denotes the jth column of $\boldsymbol{X}$.

Lemma 4. *Under the same assumptions as in Lemma 3, for any Mahalanobis-Frobenius norm induced by a positive definite matrix* $\boldsymbol{R} \in \mathbb{R}^{V_\mathrm{S} \times V_\mathrm{S}}$*, the matrix*

$$\boldsymbol{P}^* = \arg \min_{\boldsymbol{P}:\boldsymbol{PV}=\boldsymbol{W}} \frac{1}{2}\|\boldsymbol{M}^\top \boldsymbol{P}\|_{\boldsymbol{R}}^2, \tag{27}$$

has rank at most L*.*

Proof. Since $\boldsymbol{R}$ is positive definite, it has a decomposition $\boldsymbol{R} = \boldsymbol{N}^\top\boldsymbol{N}$, where $\boldsymbol{N} \in \mathbb{R}^{V_\mathrm{S} \times V_\mathrm{S}}$ is invertible. From the definition of Mahalanobis-Frobenius norm, we have that $\|\boldsymbol{M}^\top \boldsymbol{P}\|_{\boldsymbol{R}} = \|\boldsymbol{N}\boldsymbol{M}^\top \boldsymbol{P}\|_\mathrm{F}$. Since $\boldsymbol{N}$ and $\boldsymbol{M}$ are both invertible, so is $\boldsymbol{M}' := \boldsymbol{M}\boldsymbol{N}^\top$. Hence we can take Lemma 3 with $\boldsymbol{M}'$ in place of $\boldsymbol{M}$. □

A.3 Other Norms

We now prove Prop. 2. We will start by showing a counter-example to the analogous of Lemma 3, when using $\mathcal{R} = \mathcal{R}_{\ell_1}$. We choose $V_\mathrm{S} = N = K = 3, L = 2, \boldsymbol{P}^* = \boldsymbol{M} = \boldsymbol{I}_3$ and

$$\boldsymbol{V} = \boldsymbol{W} = \begin{pmatrix} -2 & 2 \\ 2 & 2 \\ 1 & 4 \end{pmatrix}. \tag{28}$$

These choices verify

$$\boldsymbol{P}^* = \arg \min_{\boldsymbol{P}:\boldsymbol{PV}=\boldsymbol{W}} \|\boldsymbol{M}^\top \boldsymbol{P}\|_1, \tag{29}$$

and $\mathrm{rank}(\boldsymbol{P}^*) = 3 > L$, which are the conditions we needed to accomplish.

Since Lemma 3 is equivalent to item 3 in Prop. 1, this proves that the analogous to item 3 in Prop. 1, when replacing $\mathcal{R}_{\ell_2}$ by $\mathcal{R}_{\ell_1}$, does not hold. The same counter-example can be used to prove that the analogous to both items 1 and 2 in Prop. 1, with $\mathcal{R} = \mathcal{R}_{\ell_1}$, do not hold.

ℓ_0-norm. The same exact counter-example above can also be used for the ℓ_0 matrix "norm," defined as the number of non-zero entries in the matrix—the solution $\boldsymbol{P}^*$ is the same as in the ℓ_1-norm case.

ℓ_∞-norm. For the ℓ_∞ matrix norm, defined as the maximum absolute value in the matrix, a very similar counter-example can be found. The only difference in this case is that the solution to

$$\boldsymbol{P}^* = \arg \min_{\boldsymbol{P}:\boldsymbol{PV}=\boldsymbol{W}} \|\boldsymbol{M}^\top \boldsymbol{P}\|_\infty, \tag{30}$$

is

$$\boldsymbol{P}^{*} = \begin{pmatrix} \frac{5}{8} & -\frac{5}{8} & \frac{1}{2} \\ -\frac{9}{40} & \frac{5}{8} & \frac{3}{10} \\ \frac{3}{8} & \frac{5}{8} & \frac{1}{2} \end{pmatrix}, \tag{31}$$

which also has rank 3.

These counter-examples have been verified with the software Mathematica, using symbolic minimization functions.