A. Supplementary Material

A.1. Minimum of (7) and (8)

Using the definition of the projective tensor norm,

$$\min_X \ell(Y,X) + \lambda \|X\|_P = \tag{25}$$

$$\min_X \ell(Y,X) + \lambda \inf_{A,Z:AZ^T=X} \sum_i \|A_i\|_a \|Z_i\|_z \tag{26}$$

Clearly, for any factorization which satisfies $X = AZ^T$ we trivially have $\ell(Y, AZ^T) = \ell(Y,X)$, so by replacing X with AZ^T and minimizing w.r.t. A and Z the equality constraint becomes redundant and we get the equivalence between the minimums of (7) and (8) (modulo the subtle difference between an infimum and a minimum).

A.2. Proof of Theorem 3

Before we present the proof, we note that the proximal operator of $\lambda\theta(\cdot)$, where $\theta(x) = \|x\|$ is any norm and $\lambda \geq 0$, is given by

$$\mathbf{prox}_{\lambda\theta}(y) = y - \Pi^{\lambda}_{\|\cdot\|^*}(y), \tag{27}$$

where $\Pi^{\lambda}_{\|\cdot\|^*}(y)$ is the projection of y onto the dual norm ball with radius λ (see e.g., Parikh & Boyd, 2013, Sec. 6.5), which is defined as

$$\Pi^{\lambda}_{\|\cdot\|^*}(y) \triangleq \arg\min_q \|y-q\|_2^2 \ \text{ s.t. } \ \|q\|^* \leq \lambda. \tag{28}$$

Let $\|\cdot\|$ be any vector norm. The proximal operator of $\theta(x) = \lambda\|x\| + \lambda_2\|x\|_2$ is the composition of the proximal operator of the l_2 norm and the proximal operator of $\|\cdot\|$, i.e., $\mathbf{prox}_\theta(y) = \mathbf{prox}_{\lambda_2\|\cdot\|_2}(\mathbf{prox}_{\lambda\|\cdot\|}(y))$.

Proof.

$$\mathbf{prox}_\theta(y) = \arg\min_{x,z:x=z} \frac{1}{2}\|y-x\|_2^2 + \lambda\|z\| + \lambda_2\|x\|_2 \tag{29}$$

This gives the Lagrangian

$$L(x,z,\gamma) = \frac{1}{2}\|y-x\|_2^2 + \lambda\|z\| + \lambda_2\|x\|_2 + \gamma^T(x-z) \tag{30}$$

Minimizing the Lagrangian w.r.t. z, we obtain

$$\min_z \lambda\|z\| - \gamma^T z = \begin{cases} 0 & \|\gamma\|^* \leq \lambda \\ -\infty & \text{else} \end{cases} \tag{31}$$

Minimizing the Lagrangian w.r.t. x, we obtain

$$\min_x \frac{1}{2}\|y-x\|_2^2 + \lambda_2\|x\|_2 + \gamma^T x = \tag{32}$$

$$\min_x \frac{1}{2}\|y-\gamma-x\|_2^2 + \lambda_2\|x\|_2 + \gamma^T y - \frac{1}{2}\|\gamma\|_2^2 = \tag{33}$$

$$\begin{cases} \frac{1}{2}\|y\|_2^2 & \|y-\gamma\|_2 \leq \lambda_2 \\ \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\left(\|y-\gamma\|_2 - \lambda_2\right)^2 & \text{else} \end{cases} \tag{34}$$

where the minimum value for x is achieved at $x = \mathbf{prox}_{\lambda_2\|\cdot\|_2}(y-\gamma)$. The relation between (32) and (33) is easily seen by expanding the quadratic terms, while the relation between (33) and (34) is given by the fact that (33) is the standard proximal operator for the l_2 norm plus terms that do not depend on x. Plugging the solution of the proximal operator of the l_2 norm (see 27, noting that the l_2 norm is self dual) into (33) gives (34). The dual of the original problem thus becomes maximizing (34) w.r.t. γ subject to $\|\gamma\|^* \leq \lambda$ with the primal-dual relation $x = \mathbf{prox}_{\lambda_2\|\cdot\|_2}(y-\gamma)$. We note that (34) is monotonically non-decreasing as $\|y-\gamma\|_2$ decreases, so the dual problem is equivalent to minimizing $\|y-\gamma\|_2$ subject to $\|\gamma\|^* \leq \lambda$, which is the dual problem of the proximal operator of the general norm (see 28), and the result follows. ■