Multiplicative updates for Nonnegative Tucker problems

Valentin Leplat, Jeremy E. Cohen, ??

1 Some notes on sparsity and normalization

Consider the toy problem

$$\underset{X\in\mathbb{R}^{m\times r}, Y\in\mathbb{R}^{r\times n}}{\text{argmin}} \; f(XY) + \mu_x\|X\|_1 + \frac{1}{2}\mu_y\|Y\|_F^2 \tag{1}$$

where f is some cost function, and $\mu_{x,y}$ are nonnegative regularization parameters. We can be more general than this using arbitrary regularizations, but let's keep it simple for now. Let us denote $\phi(X, Y)$ the cost function, and we surcharge the notation with $\phi(X, Y, \mu_x, \mu_y)$ abusively when hyperparameters may vary.

We make a few observations:

- If $\mu_y = 0$, then because $f(XY)$ is invariant to a scaling of columns (resp. rows) of X (resp. Y), we have for any couple X, Y and any $\lambda < 1$

$$\phi(\lambda X, \frac{1}{\lambda}Y) = f(XY) + \mu_x\lambda\|X\|_1 < \phi(X, Y) . \tag{2}$$

  This means that the cost can always be decreased by scaling down the columns of X, which implies that a global solution must be $X = 0$ which is however impossible. In other words this problem is degenerate and does not admit good NMF solutions, let alone sparse solutions. In practice the regularization term is will grow extremely small and the problem becomes simply argmin $_{X,Y} f(XY)$.

- Because of the scaling ambiguity, it was shown in [][Roald 2021 todo] that we can fix $\mu_x = \mu_y$ without loss of generality. Indeed, for any positive u,

$$\phi(X, Y, \mu, \mu) = f(XY) + \mu\|X\|_1 + \frac{\mu}{2}\|Y\|_F^2 = f(\frac{X}{u}uY) + \mu u\|\frac{X}{u}\|_1 + \frac{\mu}{2u^2}\|uY\|_F^2 = \phi(\tilde{X}, \tilde{Y}, \mu u, \frac{\mu}{u^2}) \tag{3}$$

  where $\tilde{X}$ and $\tilde{Y}$ are new, scaled variable. We then see that

$$\min_{X,Y} \; \phi(X, Y, \mu, \mu) = \min_{X,Y} \; \phi(\tilde{X}, \tilde{Y}, \mu u, \frac{\mu}{u^2}) = \min_{\tilde{X},\tilde{Y}} \; \phi(\tilde{X}, \tilde{Y}, \mu u, \frac{\mu}{u^2}) = \min_{X,Y} \; \phi(X, Y, \mu u, \frac{\mu}{u^2}) . \tag{4}$$

  Because any couple (μ_x, μ_y) can be written as $(\mu u, \frac{\mu}{u^2})$, setting $\mu_x = \mu_y = \mu$ does not change the minimum of the cost (but we do change the argmin). Alternatively, we can fix one of the two regularization parameters arbitrarily.

- When $\mu_x = \mu_y = \mu$, it can be shown that the solution (X^*, Y^*) must satisfy $\|X^*\|_1 = \|Y^*\|_F^2$. Indeed, fix $a = \mu\|X^*\|_1, b = \mu\|Y^*\|_F^2$, and minimize the one-dimensional cost $\phi(\lambda) := f(\frac{X}{\lambda}\lambda Y) + a\lambda + \frac{b}{\lambda^2}$ for positive lambda. It can be shown easily that $\lambda^* = (\frac{b}{a})^{1/3}$. However, if λ^* is not equal to one, it means that $\phi(X, Y)$ can be reduced by scaling and thus X^* and Y^* are not solutions. By contraposition we must have $a = b$ which yields $\|X^*\|_1 = \|Y^*\|_F^2$. This is very interesting because it means that we can balance the terms in the decomposition using penalisations on each term. In fact using the same reasoning with columnwise scaling ambiguity yields the stronger result $\|X_i^*\|_1 = \|Y_i^*\|_2^2$ for any column/row index i. This also means that we should in fact not use $\mu_x = \mu_y$ because we will not be able to decrease the norm of X without also decreasing Y which may heavily biais the fitting term $f(XY)$. Rather, a sound strategy is to fix $\mu_y = 1$ and only tune μ_x. We will then have at optimality $\mu_x\|X_i^*\|_1 = \|Y_i^*\|_2^2$ which allows to scale the ℓ_1 norm of columns of X as desired.

References