Given the nonnegative matrix factorization in the notes (ADMM Example 3).

(a) **(Written)** Assume the $\mathbf{X}$- and $\mathbf{C}$-update steps are separated, using real-valued matrix differentiation in the document "part3_minka2000realvalueddiff.pdf" posted on the course website, derive the closed-form solution $\mathbf{C}^{\star}$, $\mathbf{X}^{\star}$, and $\mathbf{R}^{\star}$ for their respective optimization problem in ADMM Example 3.

The equivalent problem is :

$$\min_{\underline{C},\underline{R},\underline{X}} \quad \| \underline{A} - \underline{X} \|_F^2 + I_+(\underline{C}) + I_+(\underline{R}) \quad \text{s.t.} \quad \underline{X} = \underline{C}\,\underline{R},$$

$$C_{ij}, R_{ij} \geq 0, \ \forall i,j \qquad I_+(\underline{C}) = \begin{cases} 0, & C_{ij} \geq 0 \ \forall i,j \\ \infty, & \text{else.} \end{cases}$$

The scaled form augmented Lagrangian :

$$\mathcal{L}_\rho(\underline{X}, \underline{C}, \underline{R}, \underline{U}) = \| \underline{A} - \underline{X} \|_F^2 + I_+(\underline{C}) + I_+(\underline{R}) + \frac{\rho}{2} \| \underline{X} - \underline{C}\,\underline{R} + \underline{U} \|_F^2$$

① For $\underline{C}$:

$$C^{k+1} = \operatorname*{argmin}_{\underline{C}} \left[\frac{\rho}{2} \| \underline{X}^k - \underline{C}\,\underline{R}^k + \underline{U}^k \|_F^2 \right]$$

$$\Rightarrow \frac{\partial}{\partial \underline{C}} \left[\frac{\rho}{2} \| \underline{X}^k - \underline{C}\,\underline{R}^k + \underline{U}^k \|_F^2 \right]$$

$$= \frac{\partial}{\partial \underline{C}} \operatorname{tr} \left[(\underline{X}^k - \underline{C}\,\underline{R}^k + \underline{U}^k)^T (\underline{X}^k - \underline{C}\,\underline{R}^k + \underline{U}^k) \right]$$

$$= \frac{\partial}{\partial \underline{C}} \operatorname{tr} \Big[\underline{X}^{k^T} \underline{X}^k - \underline{R}^{k^T} \underline{C}^T \underline{X}^k + \underline{U}^{k^T} \underline{X}^k - \underline{X}^{k^T} \underline{C}\,\underline{R}^k + \underline{R}^{k^T} \underline{C}^T \underline{C}\,\underline{R}^k - \underline{U}^{k^T} \underline{C}\,\underline{R}^k + \underline{X}^{k^T} \underline{U}^k - \underline{R}^{k^T} \underline{C}^T \underline{U}^{k^T} \underline{U}^k \Big] = 0$$

$$\Rightarrow -\underline{R}^{k^T} \underline{X}^{k^T} - \underline{R}^k \underline{X}^{k^T} + \underline{R}^k \underline{R}^{k^T} \underline{C} + \underline{R}^k \underline{R}^{k^T} \underline{C}^T - \underline{R}^k \underline{U}^{k^T} - \underline{R}^k \underline{U}^{k^T} = \underline{0}$$

$$\Rightarrow \underline{\underline{C}^{k+1} = (\underline{U}^k + \underline{X}^k)\, \underline{R}^{k^T} (\underline{R}^k \underline{R}^{k^T})^{-1}}$$

② For X^{k+1}:

$$X^{k+1} = \underset{X}{\arg\min}\left[\|A - X\|_F^2 + \frac{\rho}{2}\|X - \overbrace{(CR - U)}^{\triangleq N}\|_F^2 \right]$$

$$\Rightarrow \frac{\partial}{\partial X}\left[\operatorname{tr}\left[(A-X)^T(A-X)\right] + \frac{\rho}{2}\operatorname{tr}\left[(X-N)^T(X-N)\right]\right]$$

$$= -2(A^T - X^T) + \rho\cdot(X^T - N^T) = 0$$

$$\Rightarrow X^{k+1} = \frac{1}{1+\frac{\rho}{2}}\left[A + \frac{\rho}{2}(C^{k+1}R^k - U^k)\right]$$

③ For R^{k+1}:

$$R^{k+1} = \underset{R}{\arg\min}\left[\frac{\rho}{2}\|X^{k+1} - C^{k+1}R + U^k\|_F^2\right]$$

$$\Rightarrow \frac{\partial}{\partial R}\operatorname{tr}\left[\frac{\rho}{2}(X^{k+1} - C^{k+1}R + U^k)^T(X^{k+1} - C^{k+1}R + U^k)\right]$$

$$\equiv \frac{\partial}{\partial R}\operatorname{tr}\left[-R^T C^{k+1^T} X^{k+1} - X^{k+1^T} C^{k+1} R + R^T C^{k+1^T} C^{k+1} R - U^{k^T} C^{k+1} R - R^T C^{k+1^T} U^k\right]$$

$$= -2U^{k^T}C^{k+1} - 2X^{k+1^T}C^{k+1} + 2R^T C^{k+1^T}C^{k+1} = 0$$

$$\Rightarrow R^{k+1} = (C^{k+1^T}C^{k+1})^{-1}C^{k+1^T}(X^{k+1} + U^k)$$

④ For U^{k+1}: $U^{k+1} = U^k + U^{k+1} - C^{k+1}R^{k+1}$

$$\begin{cases} C^{k+1} = (U^k + X^k)R^{k^T}(R^kR^{k^T})^{-1} \\ X^{k+1} = \frac{1}{1+\frac{\rho}{2}}\left[A + \frac{\rho}{2}(C^{k+1}R^k - U^k)\right] \\ R^{k+1} = (C^{k+1^T}C^{k+1})^{-1}C^{k+1^T}(X^{k+1} + U^k) \\ U^{k+1}: \quad U^{k+1} = U^k + U^{k+1} - C^{k+1}R^{k+1} \end{cases}$$

(b) **(Written)** Derive the primal and dual feasibility conditions that make your implementation (see below) to work correctly.

primal feasibility condition : $\underline{X}^{*} = \underline{C}^{*}\underline{R}^{*}$

dual feasibility conditions:

with $\mathcal{L}_\rho(\underline{X},\underline{C},\underline{R},\underline{U}) = \|\underline{A}-\underline{X}\|_F^2 + I_+(\underline{C}) + I_+(\underline{R}) + \frac{\rho}{2}\|\underline{X}-\underline{C}\underline{R}+\underline{U}\|_F^2$

$$\begin{cases} \nabla_{\underline{X}} \mathcal{L}_\rho = 2(\underline{X}^{*T} - \underline{A}^T) + \rho(\underline{X}^{*T} - \underline{R}^{*T}\underline{C}^{*T} + \underline{U}^{*T}) = \underline{0} \\ \nabla_{\underline{C}} \mathcal{L}_\rho = \rho(\underline{R}^{*}\underline{R}^{*T}\underline{C}^{*T} - \underline{R}^{*}\underline{X}^{*T} - \underline{R}\,\underline{U}^{*T}) = \underline{0} \\ \nabla_{\underline{R}} \mathcal{L}_\rho = \rho(\underline{R}^{*T}\underline{C}^{*T}\underline{C}^{*} - \underline{U}^{*T}\underline{C}^{*} - \underline{X}^{*T}\underline{C}^{*}) = \underline{0} \end{cases}$$

↳ $\mathcal{L}_\rho$ is differatiable

(c)

Update $\underline{X}$, $\underline{C}$ row by row, update $\underline{R}$ col. by col.

```
function [x_opt,c_opt] = ADMM_xc(a,R,u,rho,n,P)
    cvx_begin quiet
        variable x_row(1,n) nonnegative
        variable c_row(1,P) nonnegative
        minimize (square_pos(norm((a-x_row),"fro"))+rho/2*square_pos(norm((x_row-c_row*R+u),"fro")))
    cvx_end
    x_opt=x_row;
    c_opt=c_row;
end
function [r_opt] = ADMM_r(x,C,u,P)
    cvx_begin quiet
        variable r_col(P,1) nonnegative
        minimize (square_pos(norm((x-C*r_col+u),"fro")))
    cvx_end
    r_opt=r_col;
end
```

```
parfor i=1:m
    [X(i,:), C(i,:)] = ADMM_xc(A(i,:),R,U(i,:),rho,n,P);
end
parfor j=1:n
    [R(:,j)] = ADMM_r(X(:,j),C,U(:,j),P);
end
U=U+X-C*R;
```

ρ and U update:

```
s = norm(rho*(R-Rp),"fro");
if error > 10*s
    rho = 2*rho;
    U = U/2;
elseif s > 10*error
    rho = rho/2;
    U = 2*U;
end
```

terminate condition with ε_{abs} and $\varepsilon_{rel} = 10^{-3}$:

```
e = 1e-3;
t1 = sqrt(m)*e + e*max([norm(X,"fro"),norm(C*R,"fro"),norm(A,"fro")]);
t2 = sqrt(m)*e + e*rho*norm(U,"fro");
if error < t1 && s < t2
    break
end
```

(d)

```
initial error = 17.43399
iter 01 finished, error: 12.33945
iter 02 finished, error: 5.77803
iter 03 finished, error: 3.61229
iter 04 finished, error: 2.44061
iter 05 finished, error: 1.34210
iter 06 finished, error: 0.51474
iter 07 finished, error: 0.20309
iter 08 finished, error: 0.08939
iter 09 finished, error: 0.05062
iter 10 finished, error: 0.03691
iter 11 finished, error: 0.02999
iter 12 finished, error: 0.02510
iter 13 finished, error: 0.02120
iter 14 finished, error: 0.01800
iter 15 finished, error: 0.01534
iter 16 finished, error: 0.01311
iter 17 finished, error: 0.01118
iter 18 finished, error: 0.00954
iter 19 finished, error: 0.00817
iter 20 finished, error: 0.00702
Sum of singular values of A  = 168.10947
Sum of singular values of CR = 160.98296
Sum of singular values error = 7.12651
```