The problem

Given $V \in \mathbb{R}^{m \times n}_+$, find $W$ and $H$:

minimizes $\quad F(W, H) = \frac{1}{2} \| V - WH \|_F^2 + \frac{1}{2} \lambda \ln \det \left( W^T W + \delta I \right)$ (1)

subject to $\quad W \in \mathbb{R}^{m \times r}_+$ (non-negative) (2)
$\quad H \in \mathbb{R}^{r \times n}_+$ (non-negative) (3)
$\quad \delta > 0$ (avoid log zero) (4)
$\quad r \ll m$ (low rank) (5)
$\quad h_j^T 1 \leq \alpha \quad \forall j$ (columns of $H$ inside $\alpha$-simplex) (6)

- (5,6) optional: $r$ pre-defined, $\alpha = 1$ and $W, H$ assume full rank
- $F$ non-convex ($f$ non-cvx $g$ concave)
- ways of attacking the problem:
  - cvx way: cast a cvx approximation by semidefinite relaxation
  - non-cvx way: develop/use fast first order method (local, coordinate)
    (want fast : not consider global non-cvx methods, second order methods)
(2/5) The CVX way: convex approximation

For \( \min_{W,H \geq 0} F = \frac{1}{2} \| V - WH \|_F^2 + \frac{1}{2} \lambda \ln \det W^T W \), let

\[
Z = \begin{bmatrix} W \\ H^T \end{bmatrix} \in \mathbb{R}_{+}^{(m+n) \times r} \implies ZZ^T = \begin{bmatrix} WW^T & WH \\ H^T W & H^T H \end{bmatrix} \in \mathbb{R}_{+}^{(m+n) \times (m+n)}
\]

solve \( F \) by casting a conic minimization problem on variable \( Z \):

- single variable, cvx, stable, global solution, but
- how to define non-negativity?
- such reformulation only solve the non-cvx issue of \( f = \frac{1}{2} \| V - WH \|_F^2 \)
- convexifies non-cvx \( g = \ln \det W^T W \) will go crazy
- if Big data \( \implies Z \) is Big and \( ZZ^T \) is Big\(^2\), (too expensive?)
- related to Completely Positive factorization, cp-rank … (and go crazy)
The general algorithm for the two-variable problem has the coordinate-descent-like / alternating structure

\[
\begin{align*}
\min_{W,H \geq 0} F(W, H) &= \frac{1}{2} \|V - WH\|_F^2 + \frac{1}{2} \lambda \ln \det(W^TW) \\
&= f + g
\end{align*}
\]

1. **INPUT**: \(V\)
2. **Initialization**: \(W_0, H_0\) and other parameters (if any)
3. **Loop until converge**:
   4. **Fix** \(H\), update \(W\) via \(F(W) = f(W) + g(W)\)
   5. **Fix** \(W\), update \(H\) via \(F(H) = f(H)\)

where the "update" is performed as some first order fixed-point equations.
<table>
<thead>
<tr>
<th>Problem</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconstrained, cvx ( \min_{x \in \mathbb{R}^n} f(x) )</td>
<td>Gradient Descent (1-stage) ( x_{k+1} = x_k - t_k \nabla f(x_k) )</td>
</tr>
<tr>
<td>Constrainted, cvx ( \min_{x \in Q \subset \mathbb{R}^n} f(x) )</td>
<td>Projected GD (2-stage) ( y_{k+1} = x_k - t_k \nabla f(x_k) ) ( x_{k+1} = \Pi_Q(y_{k+1}) )</td>
</tr>
<tr>
<td>Unconstrained, regularized, cvx ( \min_{x \in \mathbb{R}^n} f(x) + g(x) )</td>
<td>Proximal GD (2-stage) ( y_{k+1} = x_k - t_k \nabla f(x_k) ) ( x_{k+1} = \text{prox}<em>{\alpha g}(y</em>{k+1}) )</td>
</tr>
<tr>
<td>Unconstrained, regularized, cvx ( \min_{x \in \mathbb{R}^n} f(x) + g(x) )</td>
<td>Accelerated PGD (3-stage) [Beck-Teboulle 08] ( y_k = x_k + \frac{\lambda_{k-1}-1}{\lambda_k}(x_k - x_{k-1}) ) ( z_k = y_k - t_k \nabla f(y_k) ) ( x_{k+1} = \text{prox}_{\alpha g}(z_k) )</td>
</tr>
<tr>
<td>Unconstrained, regularized, cvx ( \min_{x \in \mathbb{R}^n} f(x) + g(x) )</td>
<td>Monotone APGD (4-stage) [Beck-Teboulle 09] ( y_k = x_k + \frac{\lambda_{k-1}-1}{\lambda_k}(z_k - x_k) + \frac{\lambda_{k-1}-1}{\lambda_k}(x_k - x_{k-1}) ) ( z_k = y_k - t_k \nabla f(y_k) ) ( v_k = \text{prox}<em>{\alpha g}(z_k) ) ( x</em>{k+1} = \begin{cases} v_k &amp; \text{if } F(v_k) \leq F(x_k) \ x_k &amp; \text{else} \end{cases} )</td>
</tr>
<tr>
<td>Unconstrained, regularized, non-cvx ( \min_{x \in \mathbb{R}^n} f(x) + g(x) )</td>
<td>MAPGD for non-cvx (6-stage) [very recent paper] ( y_k = x_k + \frac{\lambda_{k-1}-1}{\lambda_k}(z_k - x_k) + \frac{\lambda_{k-1}-1}{\lambda_k}(x_k - x_{k-1}) ) ( z_k = y_k - t_k \nabla f(y_k) ) ( w_k = \text{prox}<em>{\alpha g}(z_k) ) ( v_k = x_k - t_k \nabla f(x_k) ) ( u_k = \text{prox}</em>{\alpha g}(v_k) ) ( x_{k+1} = \begin{cases} w_k &amp; \text{if } F(w_k) \leq F(u_k) \ u_k &amp; \text{else} \end{cases} )</td>
</tr>
<tr>
<td>Constrainted, regularized, non-cvx ( \min_{x \in Q} f(x) + g(x) )</td>
<td>???</td>
</tr>
</tbody>
</table>
Apply the first order methods on

$$\min_{W,H \geq 0} F(W, H) = \frac{1}{2} \| V - WH \|_F^2 + \frac{1}{2} \lambda \ln \det(W^TW + \delta I)$$

- for proximal gradient, $g$ is assumed to be convex and $\text{prox}_g$ is assumed to be "simple", but now $g = \ln \det(W^TW + \delta I)$, it is concave
- after some calculus and algebra ($\nabla_W g = (W^\dagger)^T$, $W$ is tall matrix), we get $WA + \lambda W(W^TW + \delta I)^{-T} = B$ for solving $\text{prox}_g$, not in-expensive!
- try $x_{k+1} = \Pi_Q \left( x_k - t_k \nabla F(x_k) \right)$?
- find / construct some convex surrogate function for $g$ or try various related convex approximation (like bounded above by first order Taylor?)
- or, to be more ambitious, develops a new first order method for such problem?