(1/5) The problem

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

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

$$\text{subject to} \quad W \in \mathbb{R}_+^{m \times r} \quad \text{(non-negative)} \quad (2)$$

$$H \in \mathbb{R}_+^{r \times n} \quad \text{(non-negative)} \quad (3)$$

$$\delta > 0 \quad \text{(avoid log zero)} \quad (4)$$

$$r \ll m \quad \text{(low rank)} \quad (5)$$

$$h_j^T \mathbf{1} \leq \alpha \quad \forall j \quad \text{(columns of } H \text{ inside } \alpha\text{-simplex)} \quad (6)$$

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

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

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 $g = \ln\det W^TW$ 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)

(3/5) Non-CVX way : fast local first order method

The general algorithm for the two-variable problem has the coordinate-descent-like / alternating structure

$$\min_{W,H \geq 0} F(W,H) = \underbrace{\frac{1}{2}\|V - WH\|_F^2}_{f} + \underbrace{\frac{1}{2}\lambda \ln \det(W^T W)}_{g}$$

INPUT : V
Initialization: W_0, H_0 and other parameters (if any)
Loop until converge:
    Fix H, update W via $F(W) = f(W) + g(W)$
    Fix W, update H via $F(H) = f(H)$

where the "update" is performed as some first order fixed-point equations

The Chronicle of first order methods

Problem	Method
Unconstrainted, cvx $\min_{x\in\mathbb{R}^n} f(x)$	Gradient Descent (1-stage) $x_{k+1} = x_k - t_k \nabla f(x_k)$
Constrainted, cvx $\min_{x\in\mathcal{Q}\subset\mathbb{R}^n} f(x)$	Projected GD (2-stage) $y_{k+1} = x_k - t_k \nabla f(x_k)$ $x_{k+1} = \Pi_{\mathcal{Q}}(y_{k+1})$
Unconstrainted, regularized, cvx $\min_{x\in\mathbb{R}^n} f(x) + g(x)$	Proximal GD (2-stage) $y_{k+1} = x_k - t_k \nabla f(x_k)$ $x_{k+1} = \mathrm{prox}_{\alpha g}(y_{k+1})$
Unconstrainted, regularized, cvx $\min_{x\in\mathbb{R}^n} f(x) + g(x)$	Accelerated PGD (3-stage) [Beck-Teboulle 08] $y_k = x_k + \frac{\lambda_{k-1}-1}{\lambda_k}(x_k - x_{k-1})$ $z_k = y_k - t_k \nabla f(y_k)$ $x_{k+1} = \mathrm{prox}_{\alpha_k g}(z_k)$
Unconstrainted, regularized, cvx $\min_{x\in\mathbb{R}^n} f(x) + g(x)$	Monotone APGD (4-stage) [Beck-Teboulle 09] $y_k = x_k + \frac{\lambda_{k-1}}{\lambda_k}(z_k - x_k) + \frac{\lambda_{k-1}-1}{\lambda_k}(x_k - x_{k-1})$ $z_k = y_k - t_k \nabla f(y_k)$ $v_k = \mathrm{prox}_{\alpha_k g}(z_k)$ $x_{k+1} = \begin{cases} v_k & \text{if } F(v_k) \le F(x_k) \\ x_k & \text{else} \end{cases}$
Unconstrainted, regularized, non-cvx $\min_{x\in\mathbb{R}^n} f(x) + g(x)$	MAPGD for non-cvx (6-stage) [very recent paper] $y_k = x_k + \frac{\lambda_{k-1}}{\lambda_k}(z_k - x_k) + \frac{\lambda_{k-1}-1}{\lambda_k}(x_k - x_{k-1})$ $z_k = y_k - t_k \nabla f(y_k)$ $w_k = \mathrm{prox}_{\alpha_v g}(z_k)$ $v_k = x_k - t_k \nabla f(x_k)$ $u_k = \mathrm{prox}_{\alpha_v g}(v_k)$ $x_{k+1} = \begin{cases} w_k & \text{if } F(w_k) \le F(u_k) \\ u_k & \text{else} \end{cases}$
Constrainted, regularized, non-cvx $\min_{x\in\mathcal{Q}} f(x) + g(x)$	???

(5/5) First order methods on the very first problem

Apply the first order methods on

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

- for proximal gradient, g is assumed to be convex and 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 prox_g, not in-expensive!
- try $x_{k+1} = \Pi_{\mathcal{Q}}\Big(x_k - t_k \nabla F(x_k)\Big)$?
- find / construct some convex surrogate functon 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 ?