Supplemental Material for the paper: Online Learning in The Manifold of Low-Rank Matrices

Uri Shalit[*]**, Daphna Weinshall**
Computer Science Dept. and ICNC
The Hebrew University of Jerusalem
uri.shalit@mail.huji.ac.il
daphna@cs.huji.ac.il

Gal Chechik
Google Research and
The Gonda Brain Research Center
Bar Ilan University
gal@google.com

1 Proof of Lemma 1

We first describe an auxiliary parametrization. The tangent space at x is spanned by all the tangent vectors at 0 to the curves $\gamma : \mathbb{R} \to \mathcal{M}_k^{n,m}$ such that $\gamma(0) = x$. For any such curve, because of the rank k assumption, we may assume that for all $t \in \mathbb{R}$, there exist matrices $A(t) \in \mathbb{R}_*^{n\times k}$, $B(t) \in \mathbb{R}_*^{m\times k}$, such that $\gamma(t) = A(t)B(t)^T$. By the chain rule we have: $\dot{\gamma}(0) = \dot{A}(0)B(0)^T + A(0)\dot{B}(0)^T$. Since $x = \gamma(0) = A(0)B(0)^T$ we have:

$$T_{\mathbf{x}}\mathcal{M} = \left\{ A(0)X^T + YB(0)^T | X \in \mathbb{R}^{m\times k}, Y \in \mathbb{R}^{n\times k} \right\} \tag{1}$$

The space above is clearly a linear space. Being a tangent space to a manifold, it has the same dimension $(n+m)k - k^2$.

To prove Lemma 1, It is easy to verify that the dimension of the space defined in the Lemma is $((n+m)k - k^2$. By taking $X = MB^T + N_1B_\perp^T$ and $Y = A_\perp N_2$, it can be seen that the space above is included in $T_x\mathcal{M}_k^{n,m}$ as defined in Eq. (1) $\square$

2 Proof of theorem 1

To prove that eq. (3) of the main text is a retraction, we first show that $w_1x^\dagger w_2$ is indeed a rank k matrix. Note that $w_1 = Z_1B^T$ and likewise, $w_2 = AZ_2^T$, for some matrices $Z_1 \in \mathbb{R}_*^{n\times k}$, $Z_2 \in \mathbb{R}_*^{m\times k}$. A sufficient condition for these matrices to be of full rank is that the matrix M is of limited norm. In practice this is never a problem, as the set of matrices not of full rank is of zero measure; in applying the algorithm we have never had any issues concerning this. Thus, $R_x(\xi) = w_1x^\dagger w_2 = Z_1B^TB(B^TB)^{-1}(A^TA)^{-1}A^TAZ_2^T = Z_1Z_2^T$, which is exactly a rank-k, $n \times m$ matrix. Next we must show that $R_x(\xi)$ is a retraction, and of second order. It is obvious that $R_x(0) = x$, since the projection of the zero vector is zero, and thus ξ^S, ξ_l^P and ξ_r^P are all zero. If we expand $w_1x^\dagger w_2$ up to second order terms many terms cancel and we end up with:

$$R_x(\xi) = x + \xi^S + \xi_l^P + \xi_r^P + \xi_r^P x^\dagger \xi_l^P + O(\|\xi\|^3) = x + \xi + \xi_r^P x^\dagger \xi_l^P + O(\|\xi\|^3) \tag{2}$$

Local first order rigidity is immediately apparent. If we expand the only second order term, $\xi_r^P x^\dagger \xi_l^P$, we see that it equals $A_\perp N_2 N_1^T B_\perp^T$. We claim this term is orthogonal to the tangent space $T_x\mathcal{M}_k^{n,m}$:

$$\begin{aligned} &\left\langle \left(A_\perp N_2 N_1^T B_\perp^T\right), \left(AMB^T + AN_1^TB_\perp^T + A_\perp N_2 B^T\right) \right\rangle = \\ &\left(B_\perp N_1 N_2^T A_\perp^T AMB^T + B_\perp N_1 N_2^T A_\perp^T AN_1 B_\perp + B_\perp N_1 N_2^T A_\perp^T A_\perp N_2 B^T\right) = \\ &tr\left(B_\perp N_1 N_2^T A_\perp^T A_\perp N_2 B^T\right) = \\ &tr\left(B^T B_\perp N_1 N_2^T A_\perp^T A_\perp N_2\right) = 0 \end{aligned} \tag{3}$$

[*]also at the Gonda Brain Research Center, Bar Ilan University

Thus, the second order term cancels out if we project the second derivative of the curve defined by the retraction, as required by the second-order condition

$$P_x \left(\frac{\mathrm{d}R_x(\tau\xi)}{\mathrm{d}\tau^2}|_{\tau=0} \right) = 0 \quad \forall \xi \in T_x\mathcal{M}. \tag{4}$$

We see that the second order term is contained in the normal space. This concludes the proof that the retraction is a second order retraction.

3 Rank one pseudo-inverse update rule

For completeness we develop below the procedure for updating the pseudo-inverse of a rank-1 perturbed matrix [1], following the derivation of [2]. We wish to find a matrix G such that for a given matrix A along with its pseudo-inverse $A^\dagger$, and vectors of appropriate dimension c and d, we have:

$$\left(A + cd^T\right)^\dagger = A^\dagger + G \tag{5}$$

Algorithm 1 : Rank one pseudo-inverse update

Input: Matrices $A, A^\dagger \in \mathbb{R}_*^{n\times k}$, such that $A^\dagger$ is the pseudo-inverse of A, vectors $c \in \mathbb{R}^{n\times 1}$, $d \in \mathbb{R}^{k\times 1}$
Output: Matrix $Z^\dagger \in \mathbb{R}_*^{k\times n}$, such that $Z^\dagger$ is the pseudo-inverse of $A + cd^T$.

Compute:	matrix dimension
$v = A^\dagger c$	$k \times 1$
$\beta = 1 + d^T v$	1×1
$n = A^{\dagger T} d$	$n \times 1$
$\hat{n} = A^\dagger n$	$k \times 1$
$w = c - Av$	$n \times 1$
if $\beta \neq 0$ AND $\|w\| \neq 0$	
$G = \frac{\hat{n}}{\beta}$	$k \times 1$
$G = Gw^T$	$k \times n$
$s = \frac{\beta}{\|w\|^2 \|n\|^2 + \beta^2}$	1×1
$t = \frac{\|w\|^2}{\beta}\hat{n} + v$	$k \times 1$
$\hat{G} = s \cdot t \left(\frac{\|n\|^2}{\beta} w + n \right)^T$	$k \times n$
$G = G - \hat{G}$	$k \times n$
elseif $\beta = 0$ AND $\|w\| \neq 0$	
$G = -A^\dagger \frac{n}{\|n\|^2}$	$k \times 1$
$G = Gn^T$	$k \times 1$
$\hat{G} = v \frac{w^T}{\|w\|^2}$	$k \times n$
$G = G - \hat{G}$	$k \times n$
elseif $\beta \neq 0$ AND $\|w\| = 0$	
$G = -\frac{1}{\beta} v n^T$	$k \times n$
elseif $\beta = 0$ AND $\|w\| = 0$	
$\hat{v} = \frac{1}{\|v\|^2} v \left(v^T A^\dagger \right)$	$k \times n$
$\hat{n} = \frac{1}{\|n\|^2} \left(A^\dagger n \right) n^T$	$k \times n$
$G = \frac{v^T A^\dagger n}{\|v\|^2 \|n\|^2} v n^T - \hat{v} - \hat{n}$	$k \times n$
endif	
$Z^\dagger = A^\dagger + G$	

We use the fact that A has a full column rank to simplify slightly the algorithm of [2].

References

[1] C.D. Meyer. Generalized inversion of modified matrices. *SIAM Journal on Applied Mathematics*, 24(3):315–323, 1973.

[2] K. B. Petersen and M. S. Pedersen. The matrix cookbook, Oct. 2008. Version 20081110.