总结报告30

Trace norm and L1 norm relaxation and SDP

（2020.4.17）

一、contents

Pong, T. K., Tseng, P., Ji, S., & Ye, J. (2010). **Trace norm regularization: Reformulations, algorithms, and multi-task learning. SIAM Journal on Optimization**, 20(6), 3465–3489.

Brzyski, D., Hu, X., Goni, J., Ances, B., Randolph, T. W., & Harezlak, J. (2020). **A Sparsity Inducing Nuclear-Norm Estimator (SpINNEr) for Matrix-Variate Regression in Brain Connectivity Analysis.** 1–29.

Zhong, K., Han, M., Qiu, T., & Han, B. (2019). **Fault Diagnosis of Complex Processes Using Sparse Kernel Local Fisher Discriminant Analysis**. *IEEE Transactions on Neural Networks and Learning Systems*, *PP*, 1–11.

Bach, F., Jenatton, R., Mairal, J., & Obozinski, G. (2011). **Convex Optimization with sparsity-inducing penalties**. *Foundations and Trends in Machine Learning*, *4*(1), 1–106.

Williams, J. F. (2018). **Low-rank inducing norms with optimality interpretations**. Society for Industrial and Applied Mathematics, 69(6), 1712–1738.

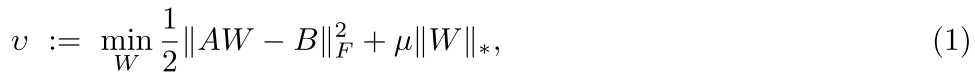
Mishra, B., Meyer, G., Bach, F., & Sepulchre, R. (2013). **Low-rank optimization with trace norm penalty.** Society for Industrial and Applied Mathematics, 23(4), 2124–2149.

二、论文1

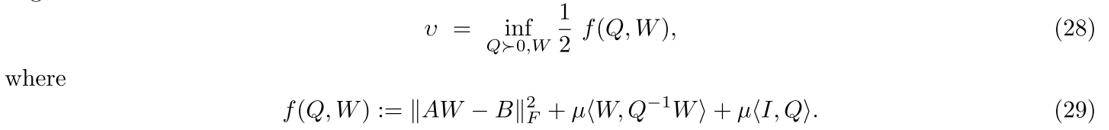
Pong, T. K., Tseng, P., Ji, S., & Ye, J. (2010). **Trace norm regularization: Reformulations, algorithms, and multi-task learning. SIAM Journal on Optimization**, 20(6), 3465–3489.

这里面给出了怎么将trace norm化成两个trace sum的形式：

即，将式（1）

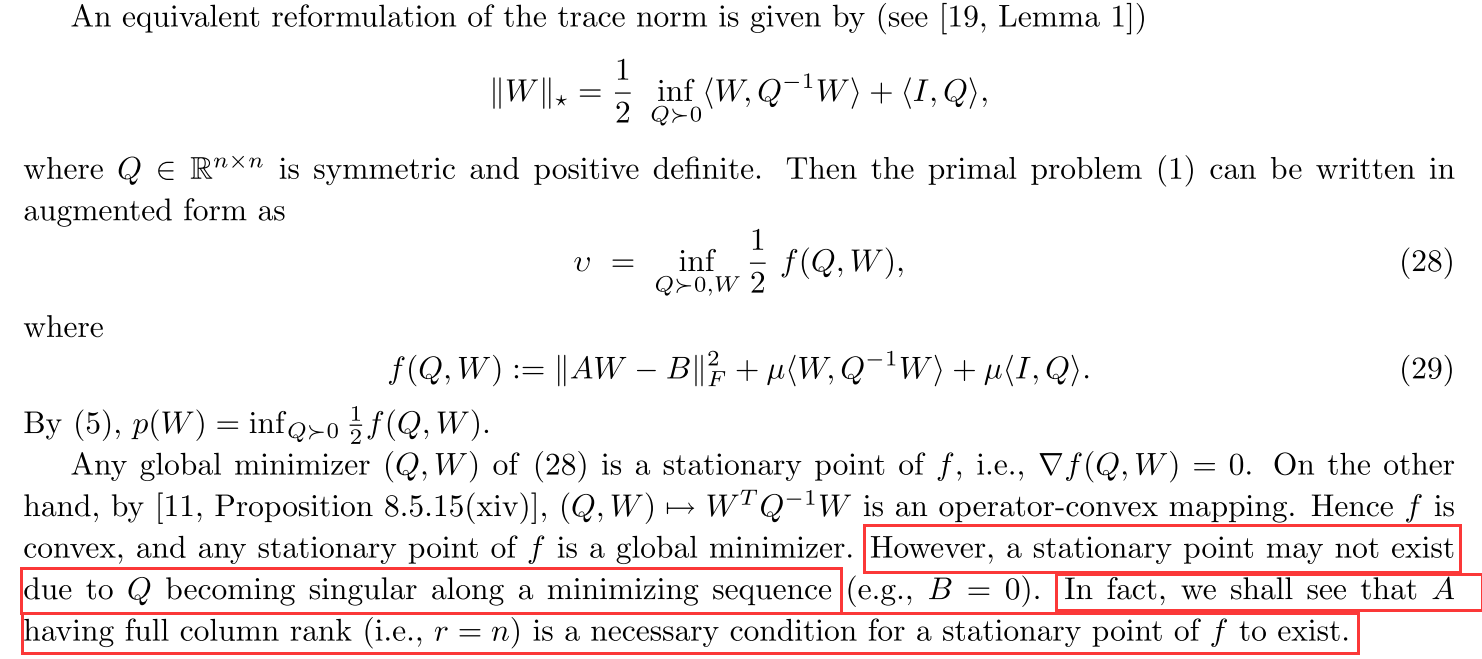


写成了



的形式。

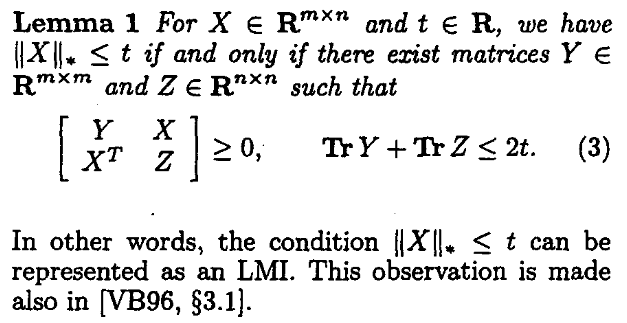
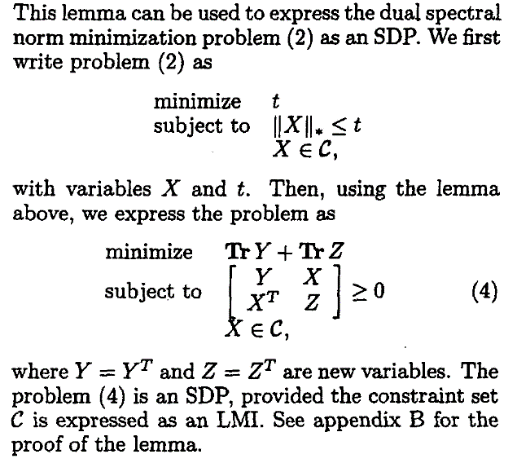
原文：



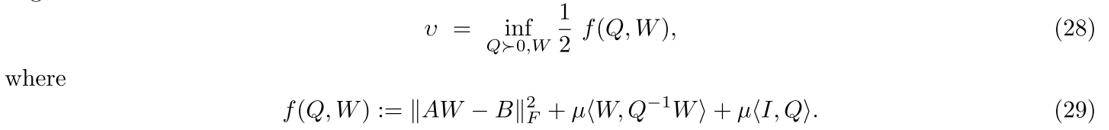
证明：

[19,Lemma 1]Fazel, M., Hindi, H., & Boyd, S. P. (2001). **A rank minimization heuristic with application to minimum order system approximation**. Proceedings of the American Control Conference, 6(2), 4734–4739.

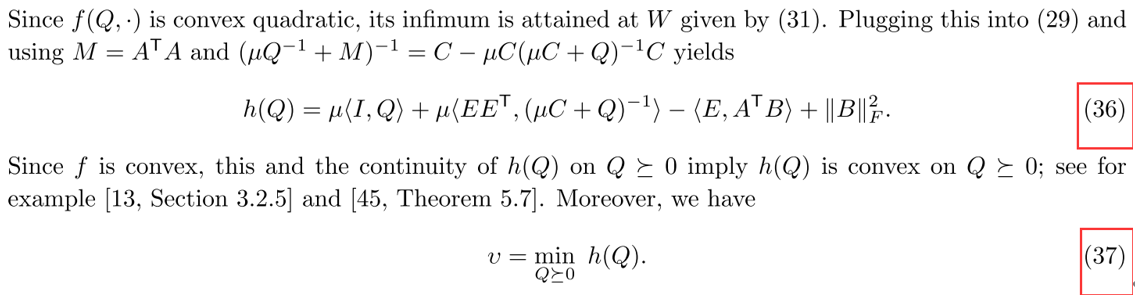
LMI——linear matrix inequality

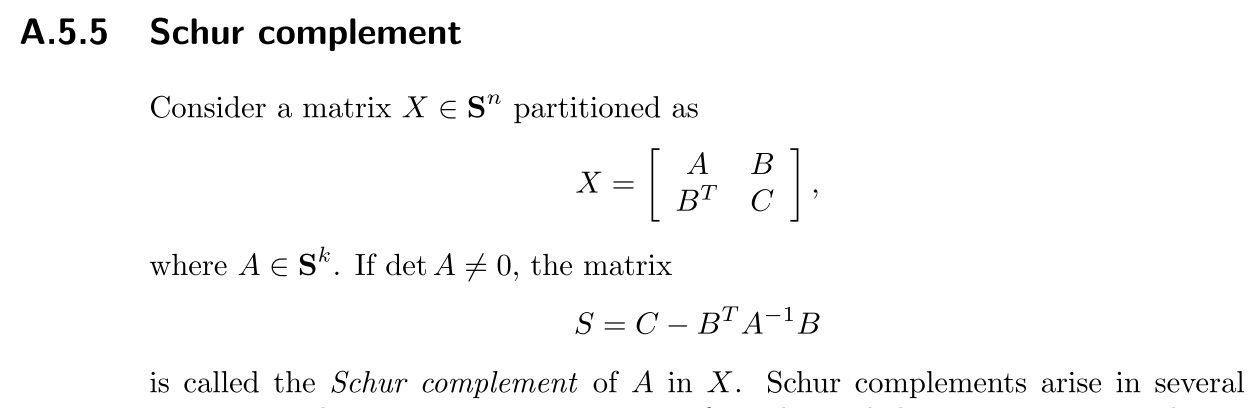
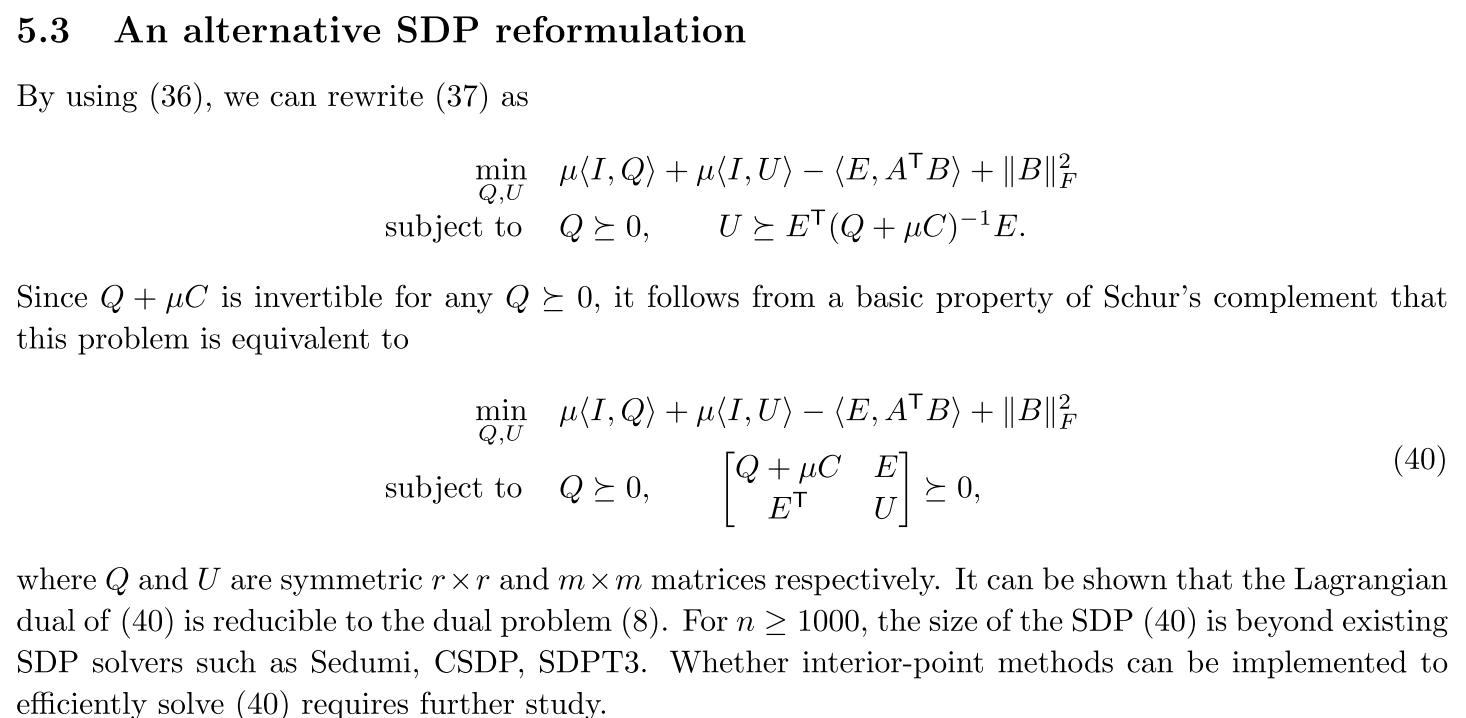
上面通过将trace norm化成两个trace sum的形式，将问题转化为



然后，



接着是利用舒尔补把trace norm转化为SDP问题：



总结：

首先，将trace norm化成两个trace sum的形式

存在的问题是：这样得到的f（Q,W）是一个凸函数，但如果Q奇异，可能不存在驻点。

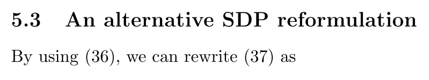
将原式写成f（Q,W）的形式。

存在的问题是：只有当f（Q,W）凸函数，可以求导，才能把求导得到的等式带入f（Q,W），进而得到h（Q）

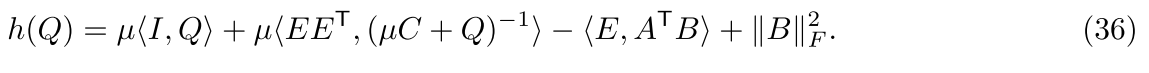
然后，对W求导带入f（Q,W），得到h（Q）。也就是将原问题转化为求min h(Q)

再利用舒尔补，得到半定规划问题。

疑问：

这个地方是怎么rewrite的。

式（36）为，

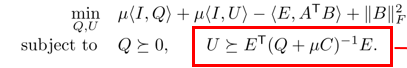


式（37）为，

为什么可以rewrite成下面的？把原来能写成半定约束形式？



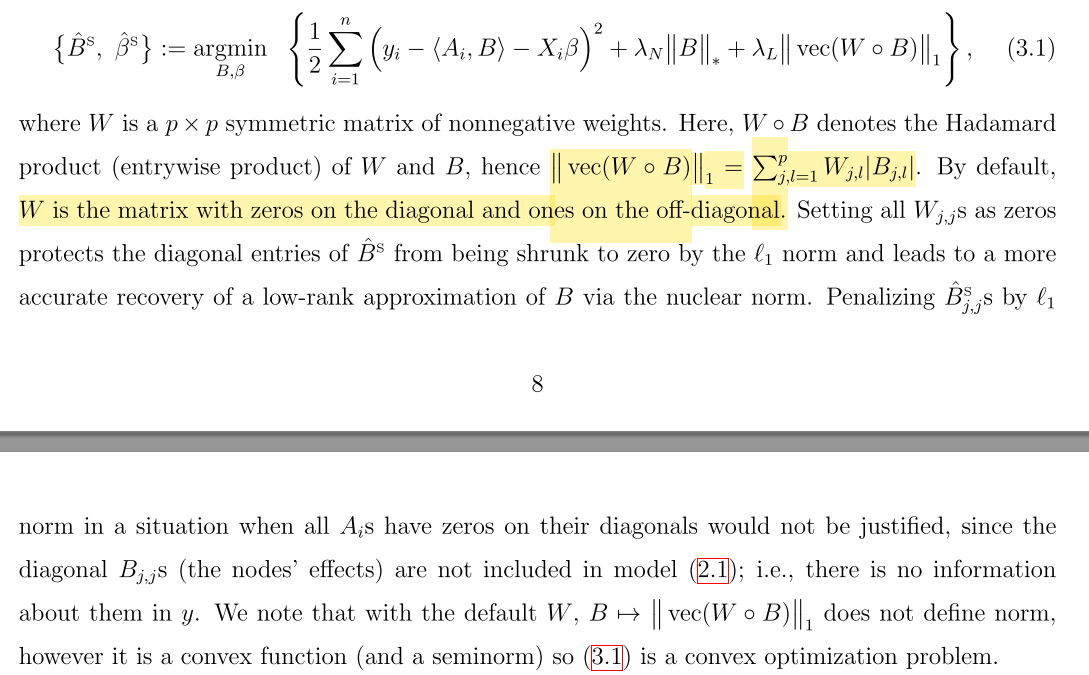
Rewrite（37）成，



三、论文2

Brzyski, D., Hu, X., Goni, J., Ances, B., Randolph, T. W., & Harezlak, J. (2020). **A Sparsity Inducing Nuclear-Norm Estimator (SpINNEr) for Matrix-Variate Regression in Brain Connectivity Analysis.** 1–29.

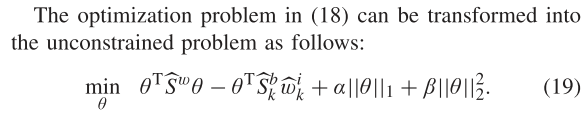
这篇论文可取之处，在于把matrix B向量化，写成vec（WB）,这样就可以把B当作向量的L1-norm来处理。



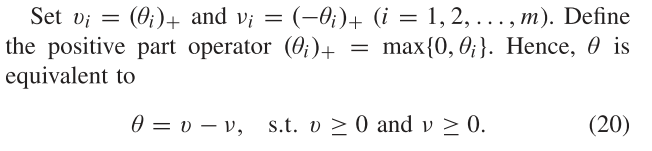
但是，他要求W是一个非负对称阵，对角元为0，非对角元为1。原因在于（这个地方不太明白，他说的是什么意思呢？要保护权值矩阵W的什么？能为我们所用吗？如果可以，接下来就可以用教研室钟凯的对向量L1-norm的松弛方法。）

四、论文3

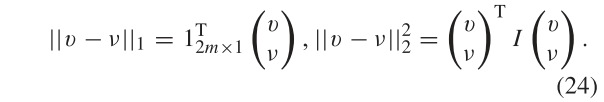
Zhong, K., Han, M., Qiu, T., & Han, B. (2019). **Fault Diagnosis of Complex Processes Using Sparse Kernel Local Fisher Discriminant Analysis**. *IEEE Transactions on Neural Networks and Learning Systems*, *PP*, 1–11.

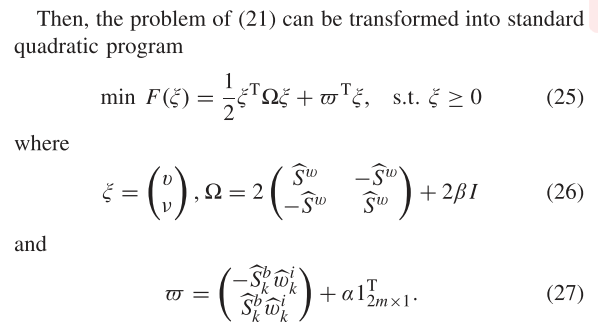


利用，



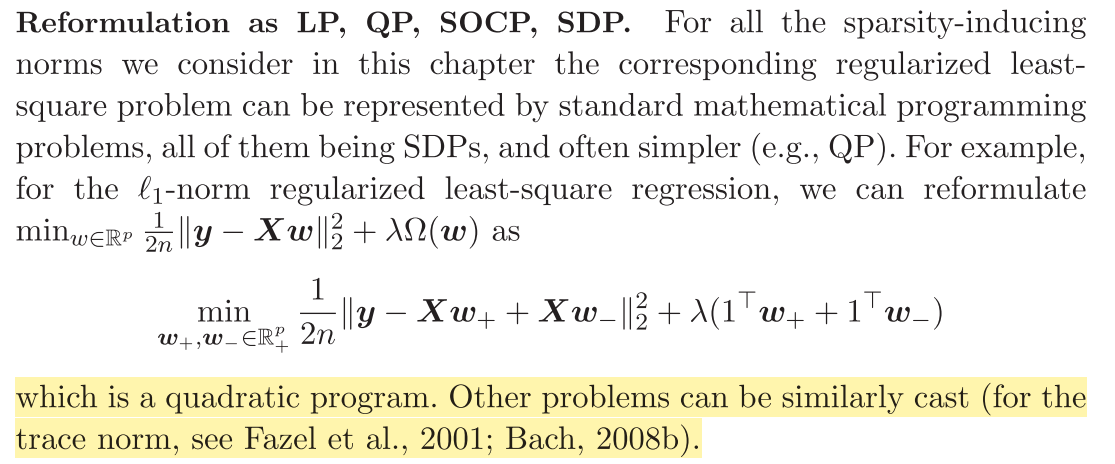
就可以把||w||的L1-norm写成下面，w是向量：





五、论文4

Bach, F., Jenatton, R., Mairal, J., & Obozinski, G. (2011). **Convex Optimization with Sparsity-Inducing Norms.** Foundations and Trends in Machine Learning, 4(1), 1–106.



这个地方我觉得和钟凯的方法挺像，是同一种方法。

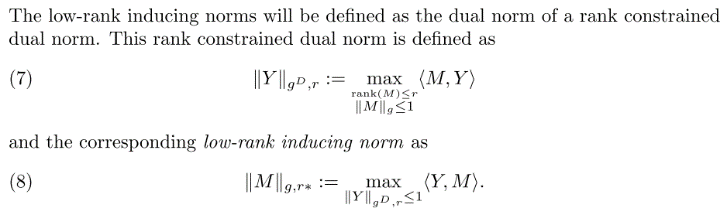
六、论文5

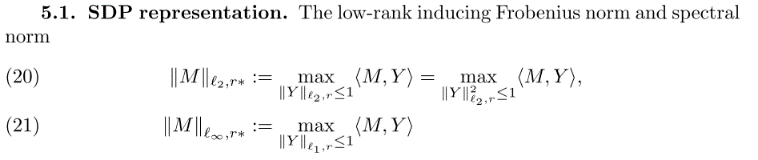
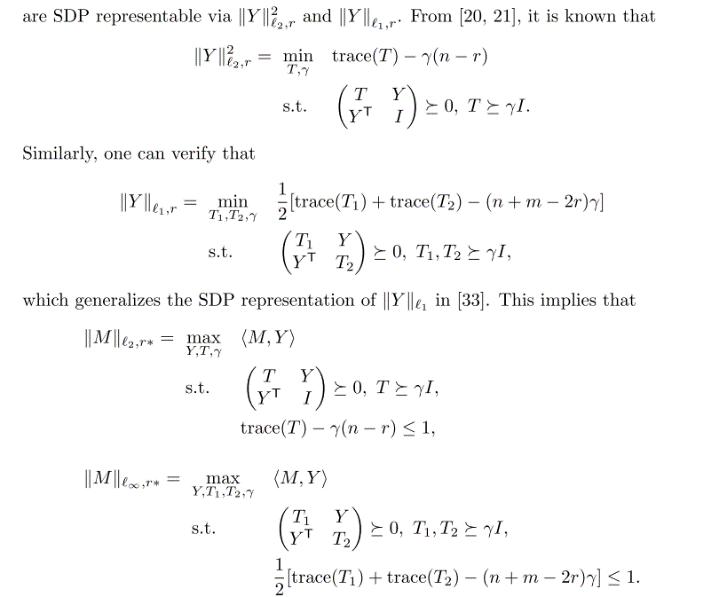
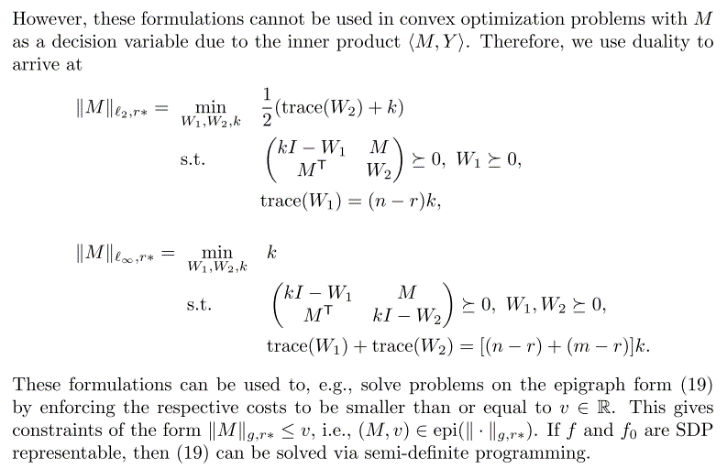
Williams, J. F. (2018). **Low-rank inducing norms with optimality interpretations**. Society for Industrial and Applied Mathematics, 69(6), 1712–1738.

（这篇和下篇论文，果然一到了数学专业的论文就看不太懂了，写的数学太严谨了，一堆符号给自己绕进去了，没看懂，等脑袋清醒再看一遍）



定义：

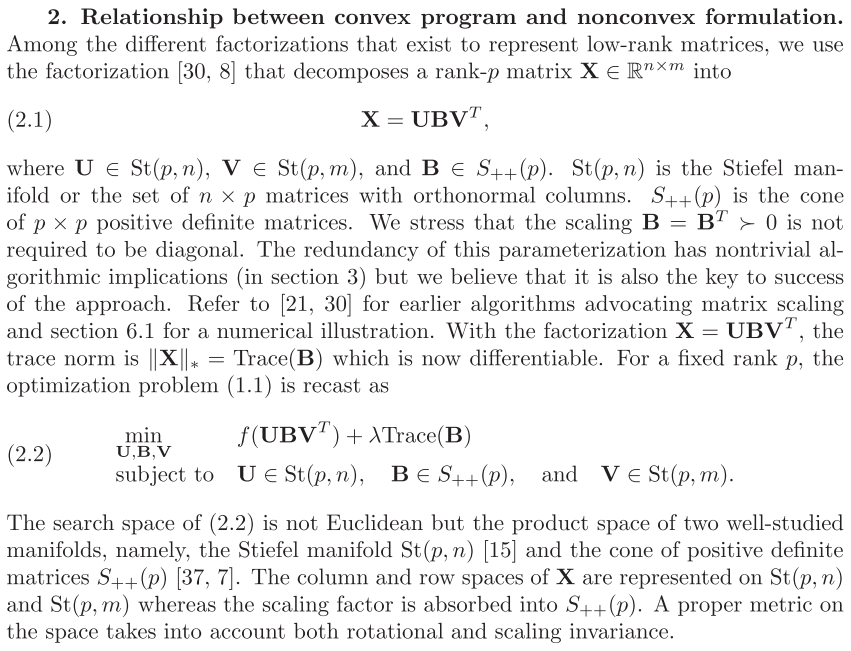


应用：

接右

七、论文6

Mishra, B., Meyer, G., Bach, F., & Sepulchre, R. (2013). **Low-rank optimization with trace norm penalty.** Society for Industrial and Applied Mathematics, 23(4), 2124–2149.



解决迹范数的。把W写成W=UBV的形式，U和V在流形上。B在一个正定阵的圆锥（cone）？

八、my idea

把原来的trace norm+L1-norm，先把trace norm写成两个trace sum 的形式，然后，在把L1 norm中的W向量化，之后应用钟凯的L1 norm relaxation。然后，再写成对偶的形式。

疑问：首先，对于钟凯的L1-norm不太理解？这种方法是不是在凸松弛？怎样凸松弛的？（我再找找有没有直接在矩阵W上L1-norm松弛的方法。）

然后，可不可以把L1-norm中的W向量化？

还有一种比较模糊的想法，就是把trace norm写成两个trace sum后，矩阵W的L1范数保持不变，或者向量化L1-norm中的W也可以。保留了一个||…||1，即L1-norm，做对偶得到对偶函数。**因为，对偶函数中有一项是1-norm的dual-norm单位球体的指示函数**。所以，在对偶范数的那个球体上，应用projection-gradient（因为projection就是向一个凸集做投影？）利用这种办法求解。

或者松弛L1-norm，保留trace-norm，对偶过去，在dual norm of trace norm的单位球体上做projection。

Re：

1. 把trace norm和L1-norm合并写成一个统一的半定规划问题，之后用原始对偶内点法。

2. 深入研究钟凯的方法，是否会产生新的比如难求导问题等。

3. 尝试下自己的方法，看下projection gradient。