More odd behavior of GD

Consider

$$\min_{X \succeq 0} F(X) = \| A(X) - y \|_2^2$$

$\uparrow$ semi definite

where $A(X) = \begin{bmatrix} \mathrm{Tr}(A_1 X) \\ \mathrm{Tr}(A_2 X) \\ \vdots \end{bmatrix}$ a linear operator

X is $n \times n$

y is $m \times 1$

Interesting case $m \ll n^2$

(so X is <u>underdetermined</u>

There are many X st $F(X) = 0$)

This means any alg has a <u>choice</u> of sol'ns — which does G.D. choose?

— is choice systematic?

Problem is equiv to

$$\min_{U \in R^{n \times d}} f(U) = \| A(UU^T) - y \|_2^2$$

(if $d < n$ there is a rank constraint, but we look at $d = n$)

· Empirical

· generate m random meas. matrices
· choose some X^*
· solve — what is soln like?
· look at

$$\frac{\| UU^T - X^* \|_F}{\| X^* \|_F}$$

this is reconstruction err; ~~it should be small~~

Small ≡ surprising, cause problem is overdetermined

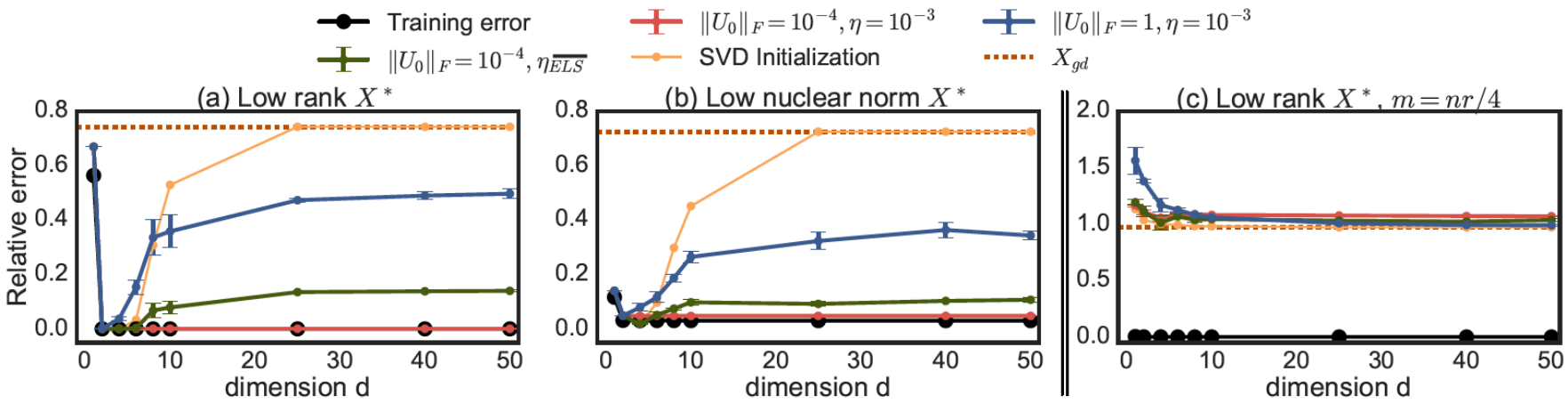

Figure 1: Reconstruction error of the global optima for 50×50 matrix reconstruction. ***(Left)*** X^* is of rank $r = 2$ and $m = 3nr$; ***(Center)*** X^* has a spectrum decaying as $O(1/k^{1.5})$ normalized to have $\|X^*\|_* = \sqrt{r}\|X^*\|_F$ for $r = 2$ and $m = 3nr$, and ***(Right)*** is a non-reconstructable setting where the number of measurements $m = nr/4$ is much smaller than the requirement to reconstruct a rank $r = 2$ matrix. The plots compare the reconstruction error of gradient descent on U for different choices initialization U_0 and step size η, including fixed step-size and exact line search clipped for stability ($\eta_{\overline{ELS}}$). Additonally, the orange dashed reference line represents the performance of X_{gd} – a rank unconstrained global optima obtained by projected gradient descent for (1) on X space, and 'SVD-Initialization' is an example of an alternate rank d global optima, where initialization U_0 is picked based on SVD of X_{gd} and gradient descent is run on factor space with small stepsize. Training error behaves similarly in all these settings (zero for $d \geq 2$) and is plotted for reference. Results are averaged across 3 random initialization and (near zero) errorbars indicate the standard deviation.

From Gunarsekar et al 2017

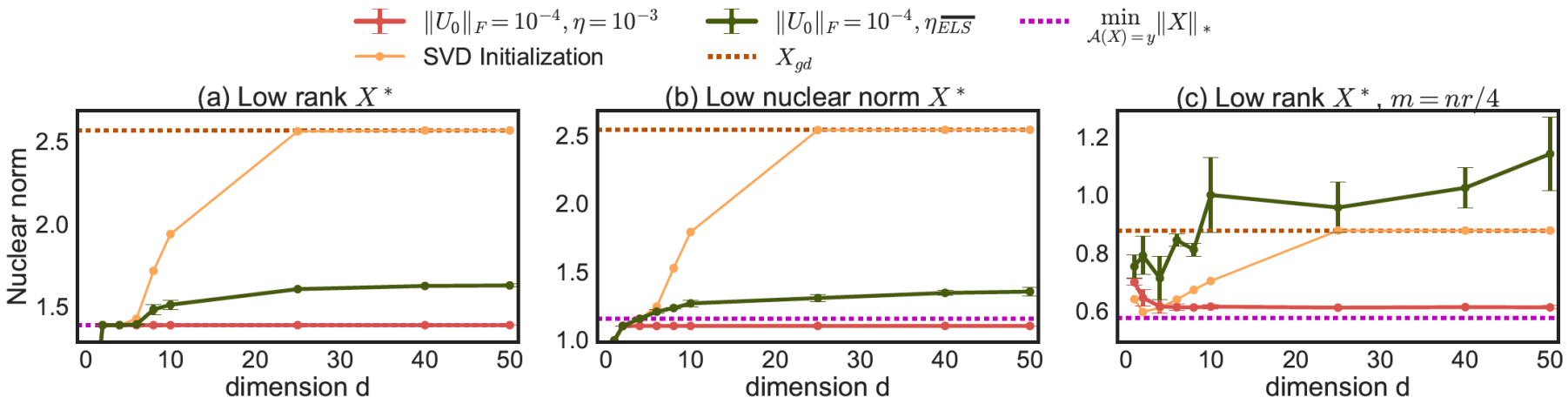

Figure 2: Nuclear norm of the solutions from Figure 1. In addition to the reference of X_{gd} from Figure 1, the magenta dashed line (almost overlapped by the plot of $\|U\|_F = 10^{-4}, \eta = 10^{-3}$) is added as a reference for the (rank unconstrained) minimum nuclear norm global optima. The error bars indicate the standard deviation across 3 random initializations. We have dropped the plot for $\|U\|_F = 1, \eta = 10^{-3}$ to reduce clutter.

From Gunarsekar et al 2017