# Bias-Variance Tradeoff in Low-Rank Approximations

# Objectives

- Introduce concept of noisy data
- Consider impact of noise on SVD
- Define bias and variance
- Use low-rank models to trade bias for variance

# Data is often contaminated by noise



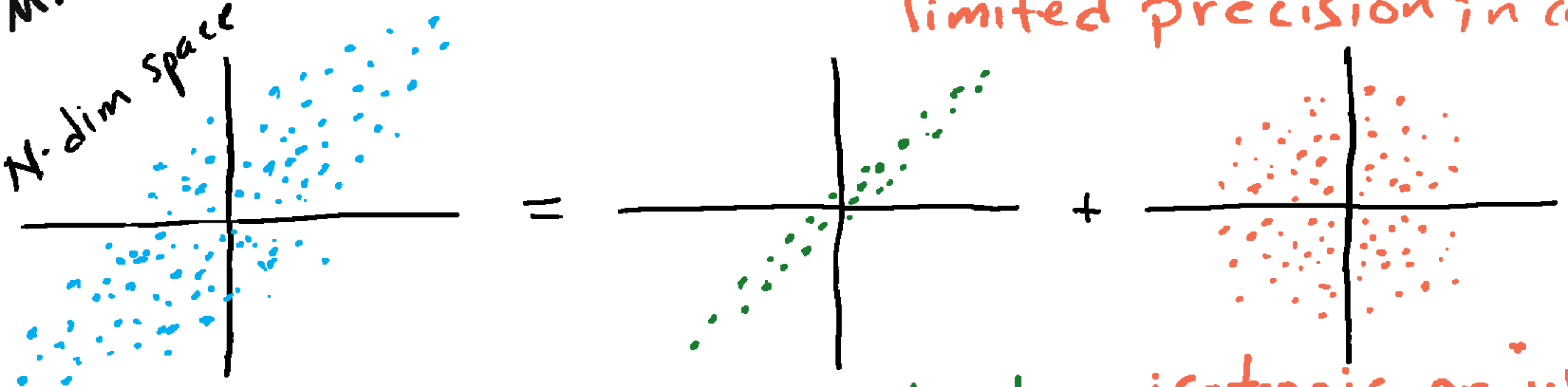
$N \times M$

$M > N$

$$\underline{A} = \underline{S} + \underline{G}$$

measured clean noise

electronics in sensing systems
enviromental static
limited precision in computers




diffuse structure

very structured

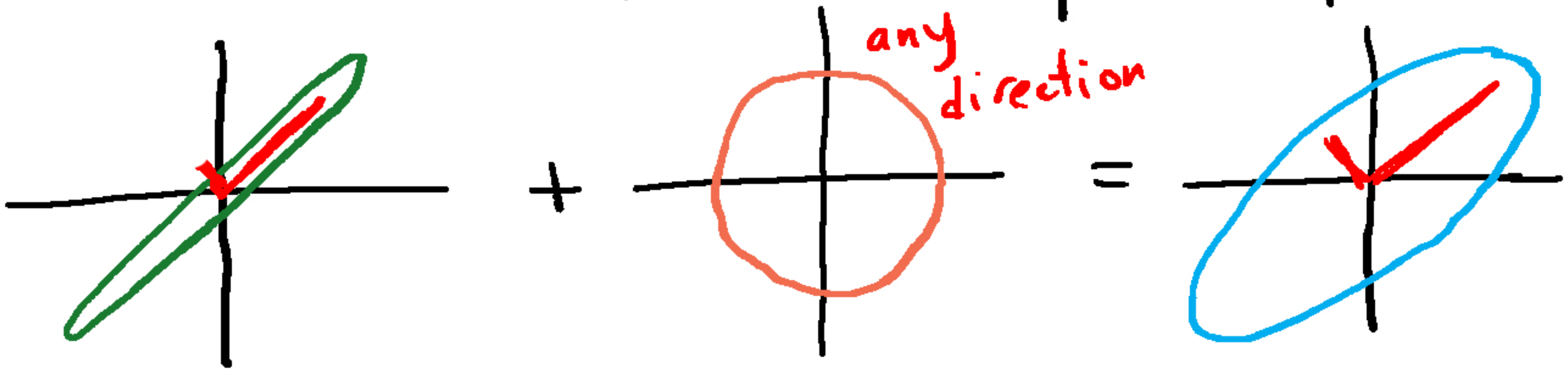
isotropic or "white" noise – no preferred direction

Sum of squared errors: $\|\underline{G}\|_F^2$

$$\|\underline{G}\|_F^2 = \sum_{i=1}^{N} M\left(\frac{1}{M}\sum_{j=1}^{M} g_{ij}^2\right) = M\sum_{i=1}^{N} var_i \overset{\text{isotropic}}{\approx} MN\sigma_g^2$$

# Singular vectors are invariant (approx) to isotropic noise

– Proof uses probability concepts




principal comp directions are unchanged ($M \to \infty$)

variance along each component (singvals) changes

$$\underline{U}\,\underline{\Sigma}_A\underline{V}^T \approx \underline{U}\,\underline{\Sigma}_S\underline{V}^T + \underline{U}\,\underline{\Sigma}_G\underline{V}^T \qquad \sigma_{A_i} \approx \sigma_{S_i} + M^{1/2}\sigma_g$$

$$\underline{\Sigma}_A = \begin{bmatrix} \sigma_{A_1} & & & 0 \\ & \sigma_{A_2} & & \\ & & \ddots & \\ 0 & & & \sigma_{A_N} \end{bmatrix}, \quad \underline{\Sigma}_S = \begin{bmatrix} \sigma_{S_1} & & & 0 \\ & \sigma_{S_2} & & \\ & & \ddots & \\ 0 & & & \sigma_{S_N} \end{bmatrix}, \quad \underline{\Sigma}_G \approx \begin{bmatrix} M^{1/2}\sigma_g & & & 0 \\ & M^{1/2}\sigma_g & & \\ & & \ddots & \\ 0 & & & M^{1/2}\sigma_g \end{bmatrix}$$

(isotropic)

$\sigma_{g_i} = M^{1/2} \cdot \mathrm{RMS}$

# Low-rank models trade bias for variance



Original: $\underline{A} = \underline{S} + \underline{G}$ Error: $\underline{A} - \underline{S}$, $\|\underline{G}\|_F^2 \approx NM\sigma_g^2$

Low rank: $\underline{\hat{A}}_r = \sum_{i=1}^{r} \sigma_{A_i} \underline{u}_i \underline{v}_i^T \approx \underline{\hat{S}}_r + \underline{\hat{G}}_r$

$$\underline{\hat{S}}_r = \sum_{i=1}^{r} \sigma_{S_i} \underline{u}_i \underline{v}_i^T \qquad \underline{\hat{G}}_r \approx \sum_{i=1}^{r} M^{1/2} \sigma_g \underline{u}_i \underline{v}_i^T$$

Bias$^2$: $b^2(r) = \|\underline{S} - \underline{\hat{S}}_r\|_F^2$

$$b^2(r) = \Big\|\sum_{i=r+1}^{N} \sigma_{S_i} \underline{u}_i \underline{v}_i^T\Big\|_F^2$$

$$= \sum_{i=r+1}^{N} \sigma_{S_i}^2$$

(notes)

sum of squared "tail" singular values

Variance: $v(r) = \|\underline{\hat{G}}_r\|_F^2$

$$v(r) = \Big\|\sum_{i=1}^{r} M^{1/2} \sigma_g \underline{u}_i \underline{v}_i^T\Big\|_F^2$$

$$= r M \sigma_g^2$$

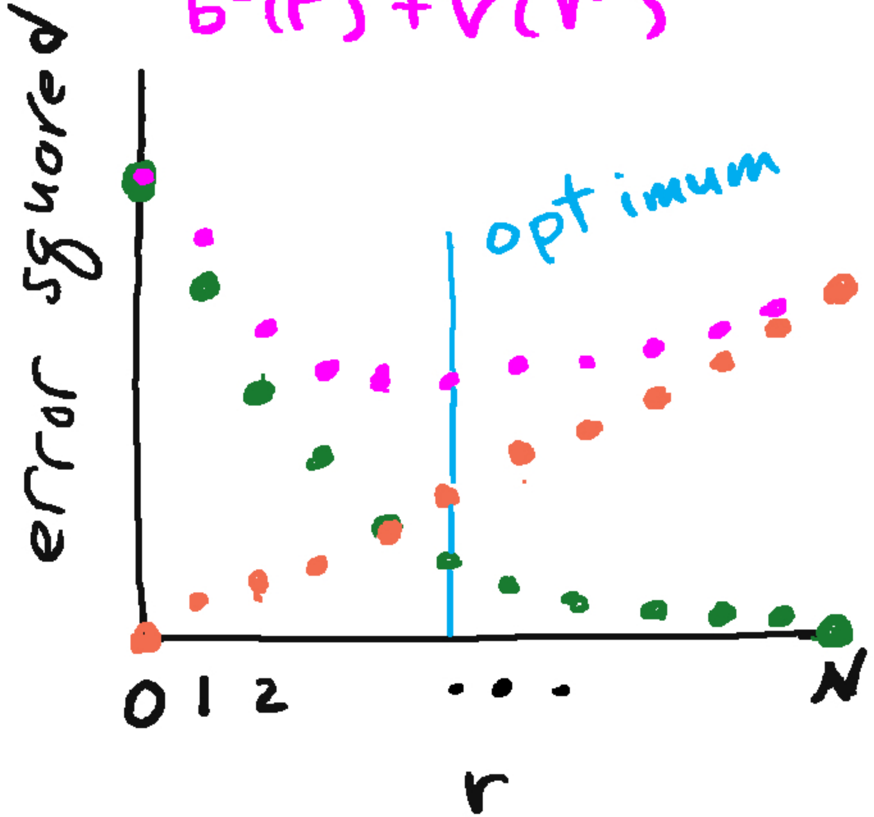
dimensions × $\frac{\text{variance}}{\text{dimension}}$

# Trading bias for variance

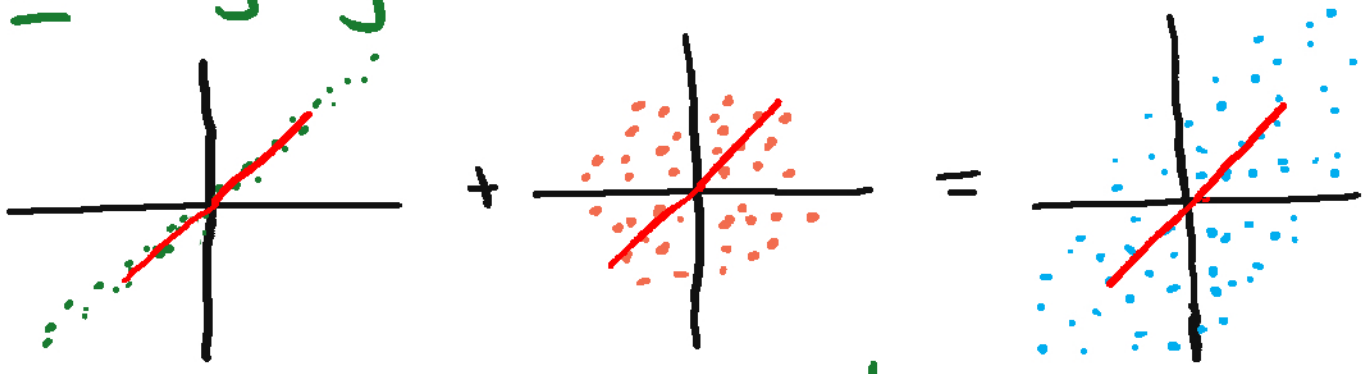
$$b^2(r) = \sum_{i=r+1}^{N} \sigma_{s_i}^2$$ decreases as $r$ increases

$$v(r) = rM\sigma_g^2$$ increases as $r$ increases

$b^2(r) + v(r)$




$\underline{S}$ highly structured $\sigma_1 \gg \sigma_2$



$\underline{S}$ weakly structured $\sigma_1 \approx \sigma_2$