Local Convergence of an Incremental Algorithm for Subspace Identification

Stephen Wright

University of Wisconsin-Madison

IPAM, January 2013

+ Laura Balzano (Michigan). (GROUSE was proposed, studied, and applied in her Ph.D. thesis, defended at UW-Madison in 2012.)

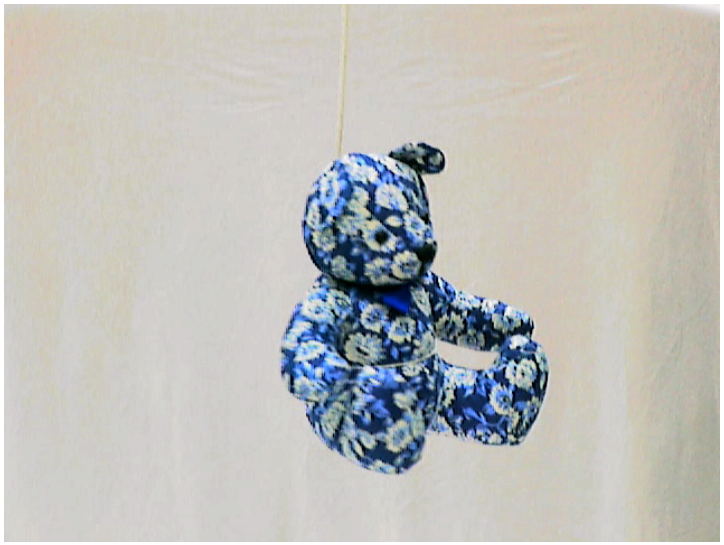
Identifying Subspaces from Partial Observations

Often we observe a certain phenomenon on a high-dimensional ambient space, but the phenomenon lies on a low-dimension subspace. Moreover, our observations may not be complete: "missing data."

Can we recover the subspace of interest?

- Matrix completion, e.g. Netflix. Observe partial rows of an $m \times n$ matrix; each row lies (roughly) in a low-d subspace of $\mathbb{R}^n$.
- Background/Foreground separation in video data.
- Mining of spatal sensor data (traffic, temperature) with high correlation between locations.
- Structure from Motion: Observe a 3-d object from different camera angles, noting the location of reference points on the object's surface on the (2-d) photo taken at each camera angle.
  - Object is solid, so some reference points are occluded in each photo. **Missing data!**
  - Matrix of reference point locations in 2-d images has rank three.
  - Range subspace reveals 3-d location of reference points.

Structure from Motion: Figures and Reconstructions



(Kennedy, Balzano, Taylor, Wright, 2012)

Euclidean Subspace Identification

- Seek subspace $S \subset \mathbb{R}^n$ of known dimension $d \ll n$.
- Know certain components $\Omega_t \subset \{1, 2, \ldots, n\}$ of vectors $v_t \in S$, $t = 1, 2, \ldots$ — the subvector $[v_t]_{\Omega_t}$.
- Assume that $\mathcal{S}$ is incoherent w.r.t. the coordinate directions.

We'll also assume for purposes of analysis that

- $v_t = \bar{U} s_t$, where $\bar{U}$ is an $n \times d$ orthonormal spanning $\mathcal{S}$ and the components of $s_t \in \mathbb{R}^d$ are i.i.d. normal with mean 0.
- Sample set Ω_t is independent for each t with $|\Omega_t| \geq q$, for some q between d and n.
- Observation subvectors $[v_t]_{\Omega_t}$ contain no noise.

Full Data: $\Omega_t \equiv \{1, 2, \ldots, n\}$: SVD (or QR)

If the vectors v_t are fully revealed — $\Omega_t \equiv \{1, 2, \ldots, n\}$ — we obtain the solution after d steps. An SVD

$$\bar{U}\bar{\Sigma}\bar{V}^T = [v_1 \,:\, v_2 \,:\, \cdots \,:\, v_d]$$

yields a spanning $n \times d$ orthonormal matrix $\bar{U}$ for $\mathcal{S}$.

Our focus is on the case of $|\Omega_t| < n$, but the analysis simplifies greatly — and gives an interesting result — in the full-data case. (More in a moment.)

Sampled Data: Batch Methods

For a fixed collection of vectors $t = 1, 2, \ldots, T$, use matrix completion: Seek $X \in \mathbb{R}^{n \times T}$ such that

$$\mathcal{A}(X) = \bar{v}, \quad \mathrm{rank}(X) = d,$$

where v is constructed from the known elements $[v_t]_{\Omega_t}$ and $\mathcal{A}$ is the corresponding location map.

Need to relax for tractability, e.g. min $\|X\|_*$ instead of imposing $\mathrm{rank}(X) = d$.

Ideally, the solution X will have

$$X = [v_1 \,:\, v_2 \,:\, \cdots \,:\, v_T].$$

A spanning matrix $\bar{U}$ can be obtained by finding the SVD of X — or of some collection of d random vectors of the form Xs, with s random.

Sampled Data: An Online / Incremental Algorithm

GROUSE (Grassmannian Rank-One Update Subspace Estimation).

- Process the v_t as a sequential stream.
- Maintain an estimate U_t (orthonormal $n \times d$) of the basis $\bar{U}$ for target subspace S;
- Simple update formula $U_t \to U_{t+1}$ when the next $(v_t)_{\Omega_t}$ is received.

Note:

- Setup is similar to incremental and stochastic gradient methods in machine learning and optimization.
- Simple rank-one update formula, akin to updates in quasi-Newton Hessian and Jacobian approximations in optimization
- Projection, so that all iterates U_t are $n \times d$ orthonormal.

One GROUSE Step

Given current estimate U_t and partial data vector $[v_t]_{\Omega_t}$, where $v_t = \bar{U} s_t$:

$$
\begin{aligned}
w_t &:= \arg\min_w \|[U_t w - v_t]_{\Omega_t}\|_2^2; \\
p_t &:= U_t w_t; \\
[r_t]_{\Omega_t} &:= [v_t - U_t w_t]_{\Omega_t}; \quad [r_t]_{\Omega_t^c} := 0; \\
\sigma_t &:= \|r_t\| \|p_t\|; \\
\text{Choose} \;\; \eta_t &> 0;
\end{aligned}
$$

$$
U_{t+1} := U_t + \left[(\cos \sigma_t \eta_t - 1) \frac{p_t}{\|p_t\|} + \sin \sigma_t \eta_t \frac{r_t}{\|r_t\|} \right] \frac{w_t^T}{\|w_t\|};
$$

We focus on the (locally acceptable) choice

$$
\eta_t = \frac{1}{\sigma_t} \arcsin \frac{\|r_t\|}{\|p_t\|}, \quad \text{which yields} \;\; \sigma_t \eta_t = \arcsin \frac{\|r_t\|}{\|p_t\|} \approx \frac{\|r_t\|}{\|p_t\|}.
$$

GROUSE Observations

With the particular step above, and assuming $\|r_t\| \ll \|p_t\|$, have

$$U_{t+1} w_t \approx U_t w_t + \frac{\|r_t\|}{\|p_t\|} \frac{r_t}{\|r_t\|} \frac{w_t^T w_t}{\|w_t\|} = p_t + r_t,$$

since $p_t = U_t w_t$. Thus

$$[U_{t+1} w_t]_{\Omega_t} \approx [p_t + r_t]_{\Omega_t} = [v_t]_{\Omega_t},$$
$$[U_{t+1} w_t]_{\Omega_t^c} \approx [p_t + r_t]_{\Omega_t^c} = [U_t w_t]_{\Omega_t},$$

where the second line follows from $[r_t]_{\Omega_t^c} = 0$. Thus

- On sample set Ω_t, $U_{t+1} w_t$ matches obervations in v_t;
- On other elements, the components of $U_{t+1} w_t$ and $U_t w_t$ are similar.
- $U_{t+1} z = U_t z$ for any z with $w_t^T z = 0$.

GROUSE Comments

The GROUSE update is essentially a project of a step along the search direction $r_t w_t^T$. Defining the inconsistency measure

$$\mathcal{E}(U_t) := \min_{w_t} \|[U_t]_{\Omega_t} w_t - [v_t]_{\Omega_t}\|_2^2,$$

we have

$$\frac{d\mathcal{E}}{dU_t} = -2 r_t w_t^T,$$

so we see that the GROUSE search direction is the negative gradient of $\mathcal{E}$.

The GROUSE update has much in common with quasi-Newton updates in optimization, in that it makes the **minimal adjustment required to match the latest observations**, while retaining a certain desired structure — orthonormality, in this case.

GROUSE Local Convergence Questions

- How to measure discrepancy between current estimate $R(U_t)$ and $\mathcal{S}$?
- Convergence behavior is obviously random, but what can we say about expected rate? Linear? If so, how fast?
- How does the analysis specialize to the full-data case?

For the first question, can use *angles between subspaces* $\phi_{t,i}$, $i = 1, 2, \ldots, d$.

$$\cos \phi_{t,i} = \sigma_i(U_t^T \bar{U}),$$

where $\sigma_i(\cdot)$ denotes the ith singular value. Define

$$\epsilon_t := \sum_{i=1}^{d} \sin^2 \phi_{t,i} = d - \sum_{i=1}^{d} \sigma_i(U_t^T \bar{U})^2 = d - \|U_t^T \bar{U}\|_F^2.$$

We seek a bound for $E[\epsilon_{t+1}|\epsilon_t]$, where the expectation is taken over the random vector s_t for which $v_t = \bar{U} s_t$.

Full-Data Case

Full-data case vastly simpler to analyze than the general case. Define

- $\theta_t := \arccos(\|p_t\|/\|v_t\|)$ is the angle between $R(U_t)$ and $\mathcal{S}$ that is revealed by the update vector v_t;
- Define $A_t := U_t^T \bar{U}$, $d \times d$, nearly orthogonal when $R(U_t) \approx \mathcal{S}$. We have $\epsilon_t = d - \|A_t\|_F^2$.

Lemma

$$\epsilon_t - \epsilon_{t+1} = \frac{\sin(\sigma_t \eta_t)\sin(2\theta_t - \sigma_t \eta_t)}{\sin^2 \theta_t} \left(1 - \frac{s_t^T A_t^T A_t A_t^T A_t s_t}{s_t^T A_t^T A_t s_t}\right),$$

The right-hand side is nonnegative for $\sigma_t \eta_t \in (0, 2\theta_t)$, and zero if $v_t \in R(U_t) = \mathcal{S}_t$ or $v_t \perp \mathcal{S}_t$.

Our favorite choice of η_t (defined above) yields $\sigma_t \eta_t = \theta_t$, which simplifies the expression above vastly:

$$\epsilon_t - \epsilon_{t+1} = 1 - \frac{s_t^T A_t^T A_t A_t^T A_t s_t}{s_t^T A_t^T A_t s_t}.$$

Dropping subscripts, we obtain

$$\frac{s^T A^T A A^T A s}{s^T A^T A s} = \frac{s^T Y \Gamma^4 Y^T s}{s^T Y \Gamma^2 Y^T s} = \frac{\tilde{s}^T \Gamma^4 \tilde{s}}{\tilde{s}^T \Gamma^2 \tilde{s}},$$

where Y is orthogonal and Γ is a diagonal matrix with elements $\cos \phi_{t,i}$ — the angles between the subspaces $R(U_t)$ and $\mathcal{S}$ defined earlier.

Lemma

Given $Q \in \mathbb{R}^{d \times d}$, suppose that $\tilde{s} \in \mathbb{R}^d$ is a random vector whose components are all i.i.d. in $\mathcal{N}(0,1)$. Then

$$E\left(\frac{\tilde{s}^T Q \tilde{s}}{\tilde{s}^T \tilde{s}}\right) = \frac{1}{d} \operatorname{trace} Q.$$

Useful, but can't quite apply it directly.

$$\frac{\tilde{s}^T\Gamma^4\tilde{s}}{\tilde{s}^T\Gamma^2\tilde{s}} = \frac{\sum \tilde{s}_i^2 \cos^4\phi_i}{\sum \tilde{s}_i^2 \cos^2\phi_i}$$
$$= \frac{\sum \tilde{s}_i^2[1 - 2\sin^2\phi_i + \sin^4\phi_i]}{\sum \tilde{s}_i^2(1 - \sin^2\phi_i)}$$
$$\approx \frac{1 - 2(\sum \tilde{s}_i^2 \sin^2\phi_i)/(\sum \tilde{s}_i^2)}{1 - (\sum \tilde{s}_i^2 \sin^2\phi_i)/(\sum \tilde{s}_i^2)} = \frac{1-2\psi}{1-\psi},$$

where $\psi := (\sum \tilde{s}_i^2 \sin^2\phi_i)/(\sum \tilde{s}_i^2)$. Two nice things about ψ:

$$E(\psi) = \frac{1}{d}\sum_{i=1}^{d} \sin^2\phi_i = \frac{1}{d}\epsilon_t, \quad 0 \le \psi \le \max_{i=1,2,\ldots,d} \sin^2\phi_i \le \epsilon_t.$$

Theorem

Suppose that $\epsilon_t \le \bar{\epsilon}$ for some $\bar{\epsilon} \in (0, 1/3)$. Then

$$E\left[\epsilon_{t+1} \,|\, \epsilon_t\right] \le \left(1 - \left(\frac{1-3\bar{\epsilon}}{1-\bar{\epsilon}}\right)\frac{1}{d}\right)\epsilon_t.$$

Full-Data: Summary

Since the sequence $\{\epsilon_t\}$ is decreasing, by the earlier lemma, we have $\epsilon_t \downarrow 0$ with probability 1 when started with $\epsilon_0 \leq \bar{\epsilon}$.

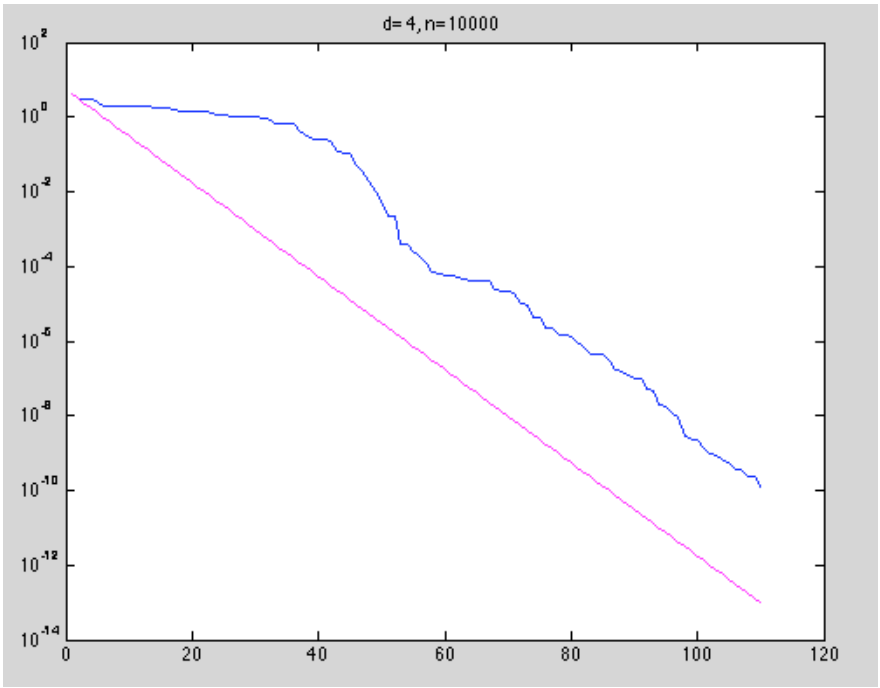
Linear convergence rate is asymptotically $1 - 1/d$.

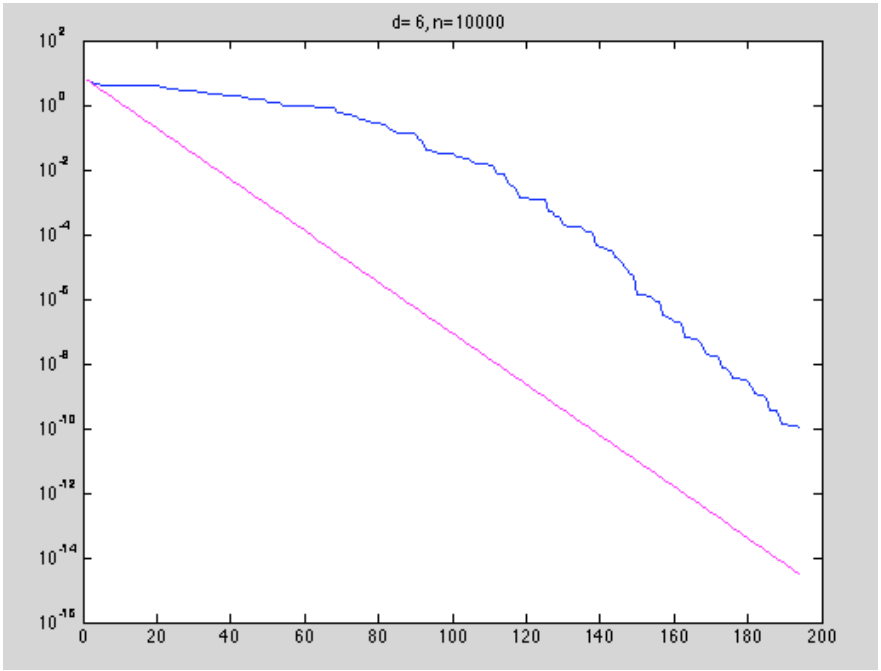
- For $d = 1$, get near-convergence in one step (thankfully!)
- Generally, in d steps (which is number of steps to get the exact solution using SVD), improvement factor is

$$(1 - 1/d)^d < \frac{1}{e}.$$

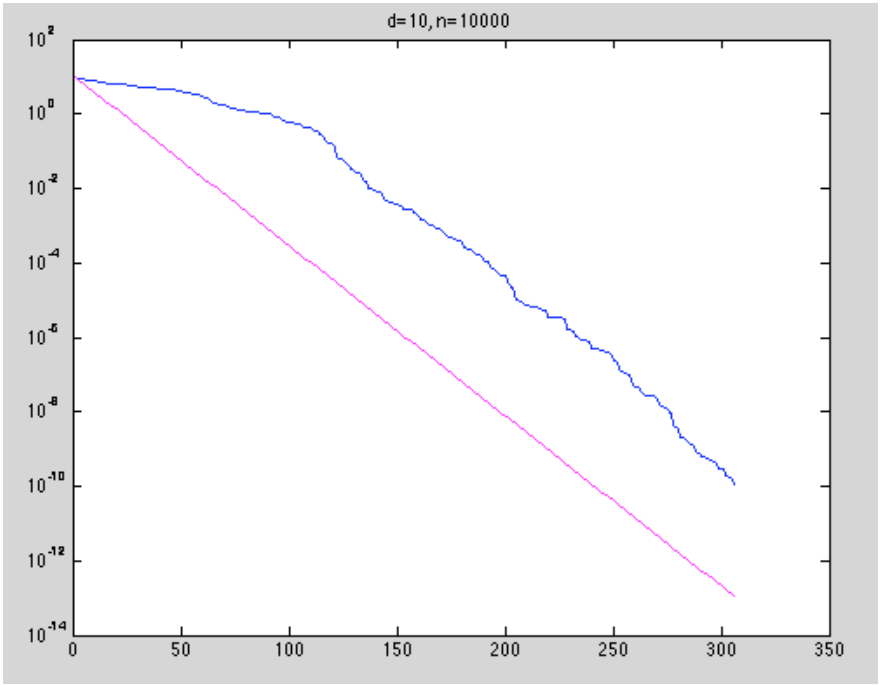
Plot some computational results for $\{\epsilon_t\}$ on a semilog plot, comparing with the curve $(1 - 1/d)^t$. $n = 10000$ and $d = 4, 6, 10, 20$.

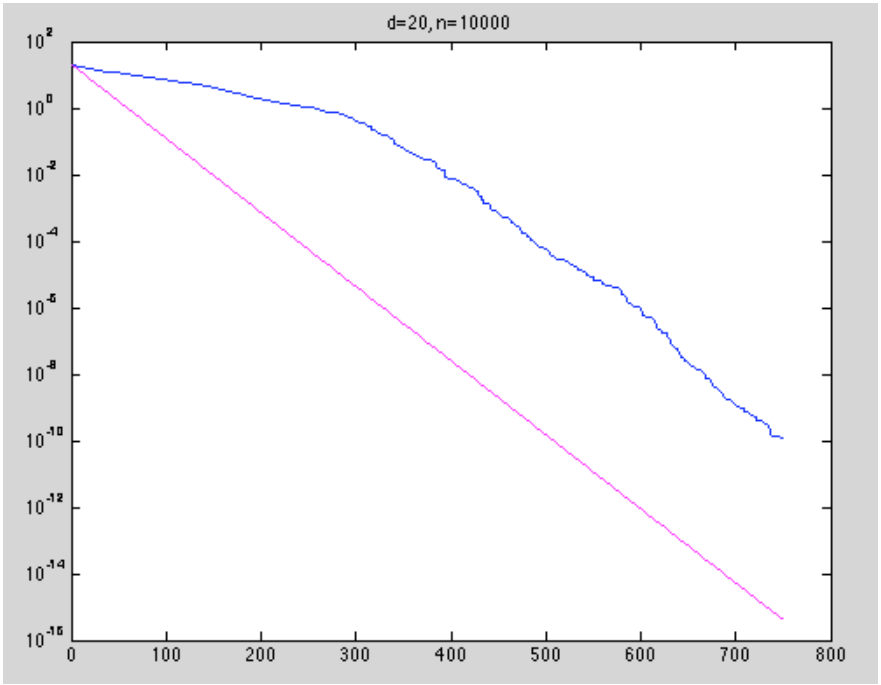
ϵ_t vs expected $(1 - 1/d)$ rate (for various d)





ϵ_t vs expected $(1 - 1/d)$ rate (for various d)





General Case: Preliminaries

Assume a regime in which ϵ_t is small.

Define coherence of $\mathcal{S}$ (w.r.t. coordinate directions) by

$$\bar{\mu} := \frac{n}{d} \max_{i=1,2,\ldots,n} \|P_{\mathcal{S}} e_i\|_2^2.$$

It's in range $[1, n/d]$, nearer the bottom if "incoherent."

Add a safeguard to GROUSE: Take the step only if

$$\sigma_i([U_t]_{\Omega_t}^T [U_t]_{\Omega_t}) \in \left[.5\frac{|\Omega_t|}{n}, 1.5\frac{|\Omega_t|}{n}\right], \quad i = 1, 2, \ldots, d,$$

i.e. the sample is big enough to capture accurately the expression of v_t in terms of the columns of U_t. Can show that this will happen w.p. $\geq .9$ if

$$|\Omega_t| \geq q \geq C_1 (\log n)^2 d\bar{\mu} \log(20d), \quad C_1 \geq \frac{64}{3}.$$

More Preliminaries

Given current measure of the distance ϵ_t from optimality, use a result from Stewart and Sun (1990) to obtain

Lemma

Suppose that $n \geq 2d$. Then there is an orthogonal $V_t \in \mathbb{R}^{d \times d}$ such that

$$\epsilon_t \leq \|\bar{U} V_t - U_t\|_F^2 \leq 2\epsilon_t.$$

Assume globally that

$$|\Omega_t| \geq q, \quad \epsilon_t \leq \frac{1}{128} \frac{q^2}{n^2 d}.$$

We can then derive several useful bounds:

$$\|r_t\| \leq \sqrt{2\epsilon_t}\|s_t\|, \quad \|p_t\| \in \left[\frac{3}{4}\|s_t\|, \frac{5}{4}\|s_t\|\right], \quad \frac{\|r_t\|^2}{\|p_t\|^2} \leq \frac{32}{9}\epsilon_t.$$

Estimate for $\epsilon_t - \epsilon_{t+1}$

Drop subscripts on r_t, w_t, p_t, σ_t, η_t. Have exactly that

$$\begin{aligned}\epsilon_t - \epsilon_{t+1} &= \|\bar{U}^T U_{t+1}\|_F^2 - \|\bar{U}^T U_t\|_F^2 \\ &= \sin^2(\sigma\eta)\left(\frac{\|\bar{U}^T r\|^2}{\|r\|^2} - \frac{\|\bar{U}^T p\|^2}{\|p\|^2}\right) + \sin(2\sigma\eta)\frac{(\bar{U}^T p)^T(\bar{U}^T r)}{\|p\|\|r\|} \\ &\geq -\sin^2(\sigma\eta) + \sin(2\sigma\eta)\frac{(\bar{U}^T p)^T(\bar{U}^T r)}{\|p\|\|r\|}.\end{aligned}$$

Our favorite choice of η yields $\sin\sigma\eta = \|r\|/\|p\|$. We can show that

$$(\bar{U}^T p)^T(\bar{U}^T r) \approx \|r\|^2.$$

Together these yield the key estimate (asymptotically exact):

$$\epsilon_t - \epsilon_{t+1} \approx \frac{\|r_t\|^2}{\|p_t\|^2}.$$

The Result

Require conditions on q and the fudge factor C_1:

$$q \geq C_1 (\log n)^2 d\bar{\mu} \log(20d), \quad C_1 \geq \frac{64}{3};$$

Also need C_1 large enough that the coherence in the residual between v_t and current subspace estimate U_t satisfies a certain (reasonable) bound w.p. $1 - \bar{\delta}$, for some $\bar{\delta} \in (0, .6)$. Then for

$$\epsilon_t \leq (8 \times 10^{-6})(.6 - \bar{\delta})^2 \frac{q^3}{n^3 d^2},$$
$$\epsilon_t \leq \frac{1}{16} \frac{d}{n} \bar{\mu},$$

we have

$$E[\epsilon_{t+1} \,|\, \epsilon_t] \leq \left(1 - (.16)(.6 - \bar{\delta}) \frac{q}{nd}\right) \epsilon_t.$$

The Result: Comments and Steps

The decrease constant it not too far from that observed in practice; we see a factor of about

$$1 - X\frac{q}{nd}$$

where X is not too much less than 1.

The threshold condition on ϵ_t is quite pessimistic, however. Linear convergence behavior is seen at much higher values.

18 pages (SIAM format) of technical analysis. We highlight the main tools and key inequalities.

Steps

1. Tightening of (deterministic) bound on $\epsilon_t - \epsilon_{t+1}$:

$$\epsilon_{t+1} \leq \epsilon_t - \frac{\|r_t\|^2}{\|p_t\|^2} + 55\sqrt{\frac{n}{q}}\epsilon_t^{3/2}.$$

If we can find a lower bound on $\|r_t\|^2/\|w_t\|^2$ as a multiple of ϵ_t, the last term becomes lower-order and we can get linear decrease, for small ϵ_t.

2.

$$\frac{\|r_t\|^2}{\|p_t\|^2} \geq \frac{16}{25}\frac{\|r_t\|^2}{\|s_t\|^2},$$

by the GROUSE safeguard (which holds for at least 90% of the iterates).

3. Use a 2010 result below: high-probability lower bound on $\|r_t\|^2$ in terms of $\|P_{N(U_t^T)}v_t\|_2^2$. **The factor** is close to $|\Omega_t|/n$ in practice, but we pay a price for coherence and for the $1-\delta$ guarantee.

(Here $\mu(.)$ denote coherence measures, which are close to 1 when the rows of the argument have similar weight, closer to n or n/d otherwise.)

Lemma (Balzano, Recht, Nowak (2010))

Let $\delta > 0$ be given, and suppose that

$$|\Omega_t| > \frac{8}{3} d\mu(U_t) \log\left(\frac{2d}{\delta}\right).$$

Then with probability at least $1 - 3\delta$, we have

$$\|r_t\|_2^2 \geq \left(\frac{|\Omega_t|(1-\xi_t) - d\mu(U_t)\frac{(1+\beta_t)^2}{1-\gamma_t}}{n}\right) \|P_{N(U_t^T)} v_t\|_2^2,$$

where

$$\xi_t := \sqrt{\frac{2\mu(P_{N(U_t^T)} v_t)^2}{|\Omega_t|} \log\left(\frac{1}{\delta}\right)}, \quad \beta_t := \sqrt{2\mu(P_{N(U_t^T)} v_t) \log\left(\frac{1}{\delta}\right)},$$

$$\gamma_t := \sqrt{\frac{8d\mu(U_t)}{3|\Omega_t|} \log\left(\frac{2d}{\delta}\right)}.$$

Steps

4. Set $\delta = .1$. We observe computationally that the error identified by the latest sample — $P_{N(U_t^T)} v_t$ — is incoherent with respect to coordinate directions. (It seems to grow like $\log n$.) We find that **the factor** is bounded below by $q/2$ when this quantity satisifies the following:

$$\mu(P_{N(U_t^T)} v_t) \leq \log n \left[\frac{.045}{\log 10} C_1 d \mu(\bar{U}) \log(20d) \right]^{1/2}$$

$$\mu(P_{N(U_t^T)} v_t) \leq (\log n)^2 \left[\frac{.05}{8 \log 10} C_1 \log(20d) \right].$$

That is, we have w.p. at least .7 that

$$\|r_t\|_2^2 \geq \frac{q}{2} \|P_{N(U_t^T)} v_t\|_2^2.$$

We assume that C_1 is chosen large enough that these bounds are satisfied w.p. at least $1 - \bar{\delta}$ for some $\bar{\delta} \in (0, .6)$.

Steps

5. Defining θ_t as the angle between v_t and the subspace $R(U_t)$, we have

$$\frac{\|P_{N(U_t^T)} v_t\|^2}{\|v_t\|^2} = \sin^2 \theta_t.$$

6. The high-probability bound now gives two cases:

$$\epsilon_{t+1} \leq \epsilon_t - .32\frac{q}{n}\sin^2\theta_t + 55\sqrt{\frac{n}{q}}\epsilon_t^{3/2}, \text{ w.p. } .6 - \bar{\delta},$$

$$\epsilon_{t+1} \leq \epsilon_t + 55\sqrt{\frac{n}{q}}\epsilon_t^{3/2}, \qquad \text{otherwise.}$$

7. Can show uisng the technical Lemma defined earlier, can show that when $v_t = \bar{U} s_t$ with components of s_t i.i.d $\mathcal{N}(0,1)$, then

$$E(\sin^2\theta_t) = \frac{1}{d}\epsilon_t.$$

The Result follows by combining all these arguments.

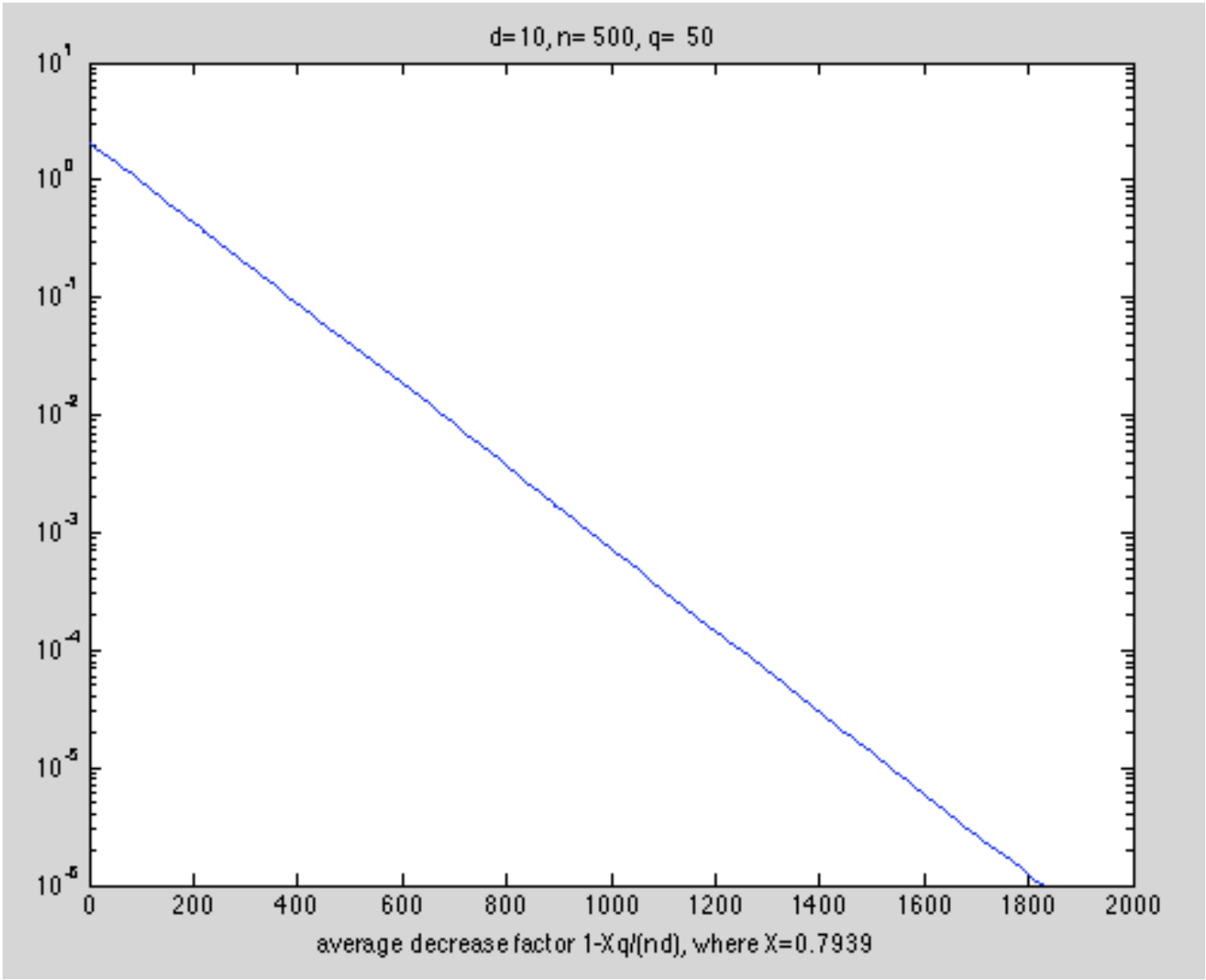
Computations for GROUSE with Sampling

- Choose U_0 so that ϵ_0 is between 1 and 4.
- Stop when $\epsilon_t \leq 10^{-6}$.
- Calculate average convergence rate: value X such that

$$\epsilon_N \approx \epsilon_0 \left(1 - X\frac{q}{nd}\right)^N.$$

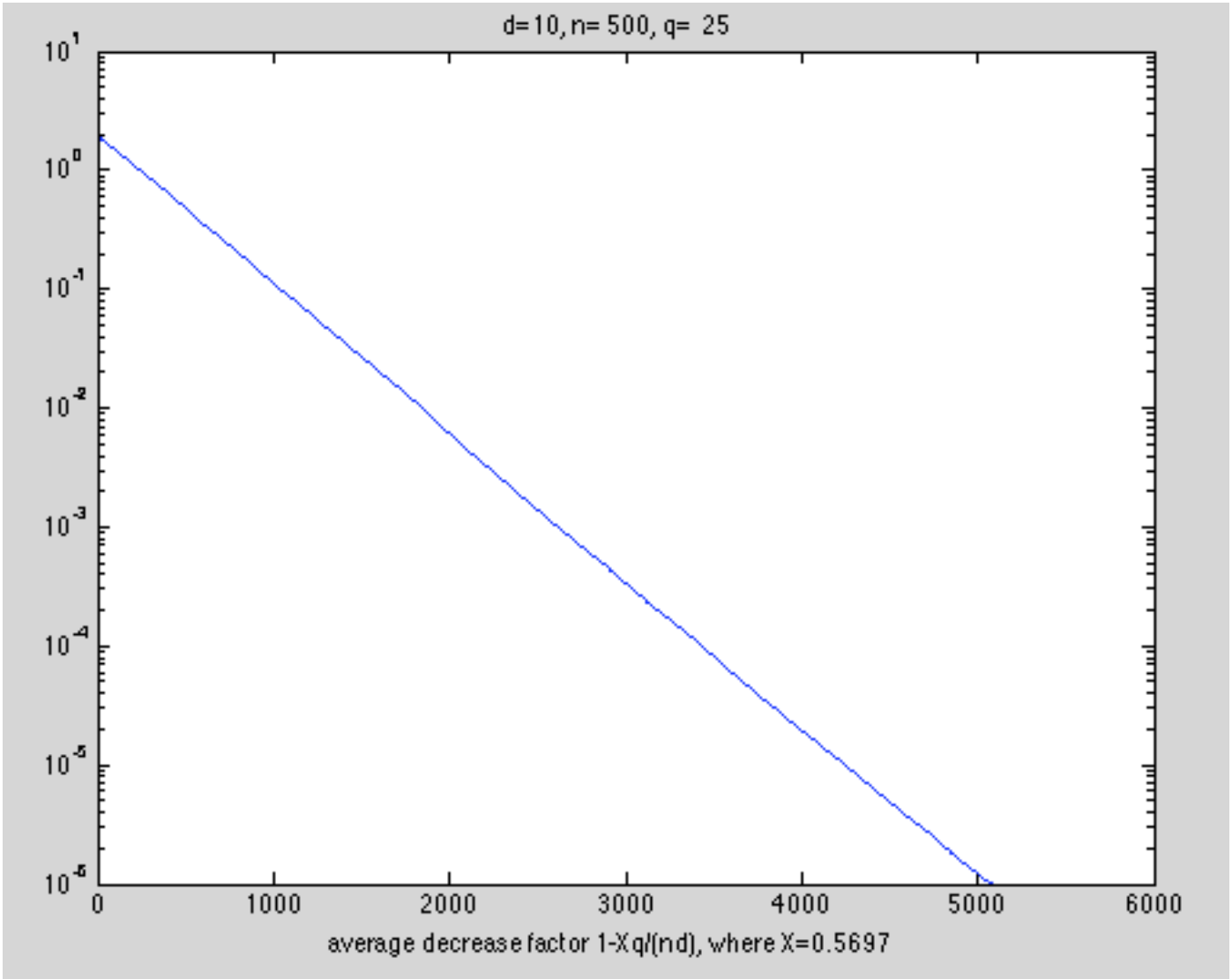
We find that X is not too much less than 1!

ϵ_t, for $n = 500$, $d = 10$, $q = 50$.



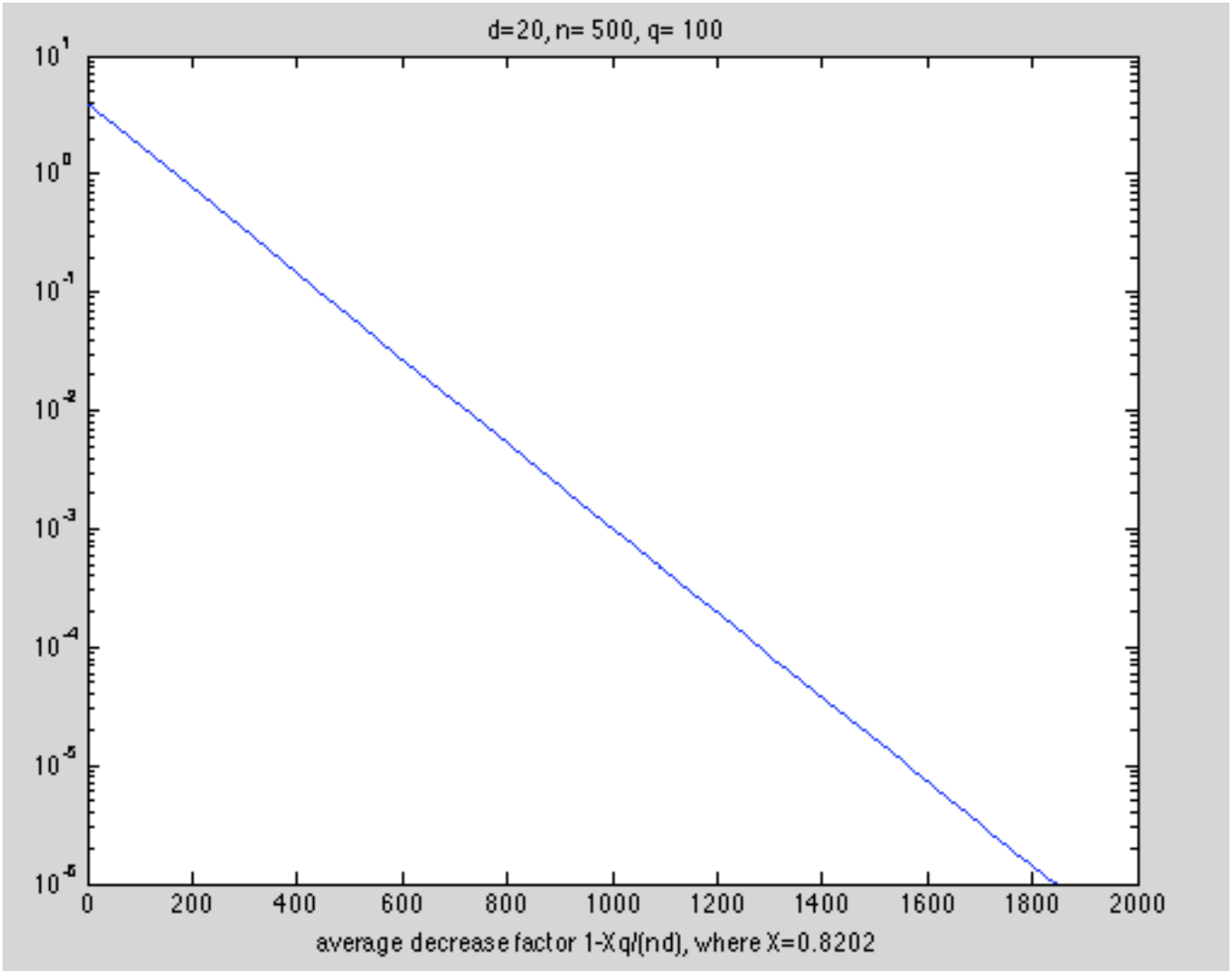
Average decrease factor $\approx 1 - .79 * q/(nd)$

ϵ_t, for $n = 500$, $d = 10$, $q = 25$.



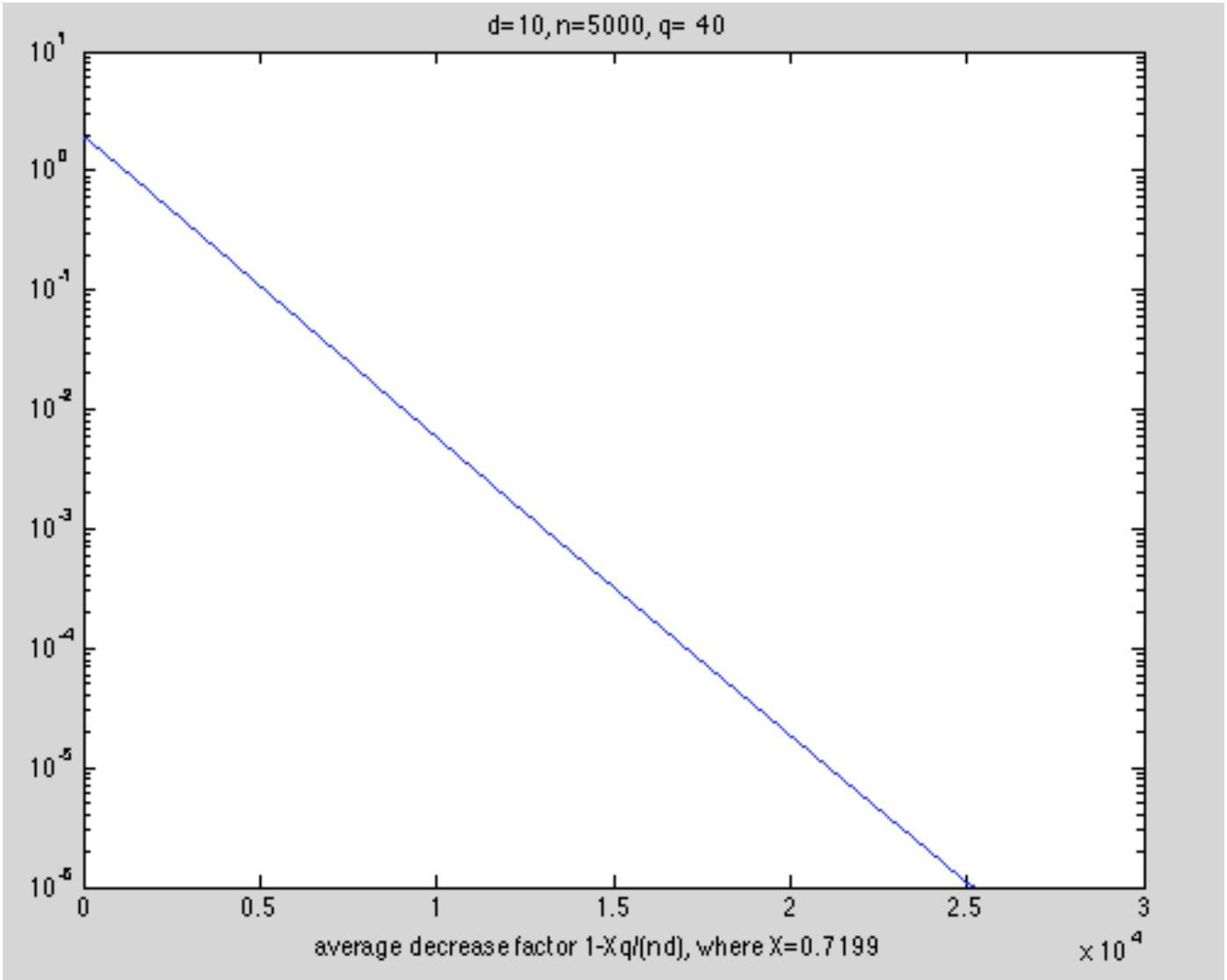
Average decrease factor $\approx 1 - .57 * q/(nd)$

ϵ_t, for $n = 500$, $d = 20$, $q = 100$.



Average decrease factor $\approx 1 - .82 * q/(nd)$

ϵ_t, for $n = 5000$, $d = 10$, $q = 40$.



Average decrease factor $\approx 1 - .72 * q/(nd)$

SVD Approaches for the General Case

A naive batch SVD approach, following the successful approach for full data, would be to assemble all the partial $[v_t]_{\Omega_t}$ into an $n \times T$ matrix, filling out with zeros, and take the estimate U_T to be the leading d singular values.

This gives terrible results — the zeros confuse it.

An incremental version, in which we update U_t by adding the column v_t (filled out with zeros), and taking the leading d singular vectors of the resulting matrix, is similarly bad.

Incremental SVD, done right: iSVD

Given U_t and $[v_t]_{\Omega_t}$:

- Compute w_t as in GROUSE:

$$w_t := \arg\min_w \|[U_t w - v_t]_{\Omega_t}\|_2^2.$$

- Use w_t to impute the unknown elements $(v_t)_{\Omega_t^C}$, and fill out v_t with these estimates:

$$\tilde{v}_t := \begin{bmatrix} [v_t]_{\Omega_t} \\ [U_t]_{\Omega_t^c} w_t \end{bmatrix}.$$

- Append $\tilde{v}_t$ to U_t and take the SVD of the resulting $n \times (d+1)$ matrix $[U_t \ : \ \tilde{v}_t]$;
- Define U_{t+1} to be the leading d singular vectors. (Discard the singular vector that corresponds to the smallest singular value of the augmented matrix.)

iSVD and GROUSE

iSVD and GROUSE seem similar:

- Both compute and use w_t to extract the missing information from U_t and $[v_t]_{\Omega_t}$.
- Both generate a sequence $\{U_t\}$ of orthonormal estimates of $\mathcal{S}$.
- Both ostensibly use no information before U_t.
- Neither has different confidence for different subspaces of the target subspace $\mathcal{S}$; both maintain a "flat" approximation.

Indeed, can show that iSVD and GROUSE are identical for certain choices of the parameter η_t.

The choice of η_t is *not* the same as the "optimal" choice in GROUSE, but it works farily well in practice.

Relating iSVD and GROUSE

Theorem

Suppose we have the same U_t and $[v_t]_{\Omega_t}$ at the t-th iterations of iSVD and GROUSE. Then there exists $\eta_t > 0$ in GROUSE such that the next iterates U_{t+1} of both algorithms are identical, to within an orthogonal transformation by the $d \times d$ matrix

$$W_t := \left[\frac{w_t}{\|w_t\|} \mid Z_t \right],$$

where Z_t is a $d \times (d-1)$ orthonormal matrix whose columns span $N(w_t^T)$.

GROUSE and iSVD: Details

The precise values for which GROUSE and iSVD are identical are:

$$\lambda = \frac{1}{2}\left[(\|w_t\|^2 + \|r_t\|^2 + 1) + \sqrt{(\|w_t\|^2 + \|r_t\|^2 + 1)^2 - 4\|r_t\|^2}\right]$$

$$\beta = \frac{\|r_t\|^2\|w_t\|^2}{\|r_t\|^2\|w_t\|^2 + (\lambda - \|r_t\|^2)^2}$$

$$\alpha = \frac{\|r_t\|(\lambda - \|r_t\|^2)}{\|r_t\|^2\|w_t\|^2 + (\lambda - \|r_t\|^2)^2}$$

$$\eta_t = \frac{1}{\sigma_t}\arcsin\beta.$$

FIN