Gene Golub Day

Memorial Workshop HKBU

# SVD – the Swiss Army Knife of Linear Algebra

Walter Gander

February 7, 2015

# Swiss Army Knife?

Dianne P. O'Leary

Matrix Factorizations for Information Retrieval (2006)

https://www.cs.umd.edu/users/oleary/a600/yahoo.pdf

# Slide from Dianne's Talk on Information Retrieval

## Uses of the SVD

This is the Swiss Army knife of matrix decompositions.

- solving ill-conditioned least squares problems
- solving discretized ill-posed problems
- solving linear systems
- determining the rank of a matrix
- determining a low-rank approximation to a matrix
- ...

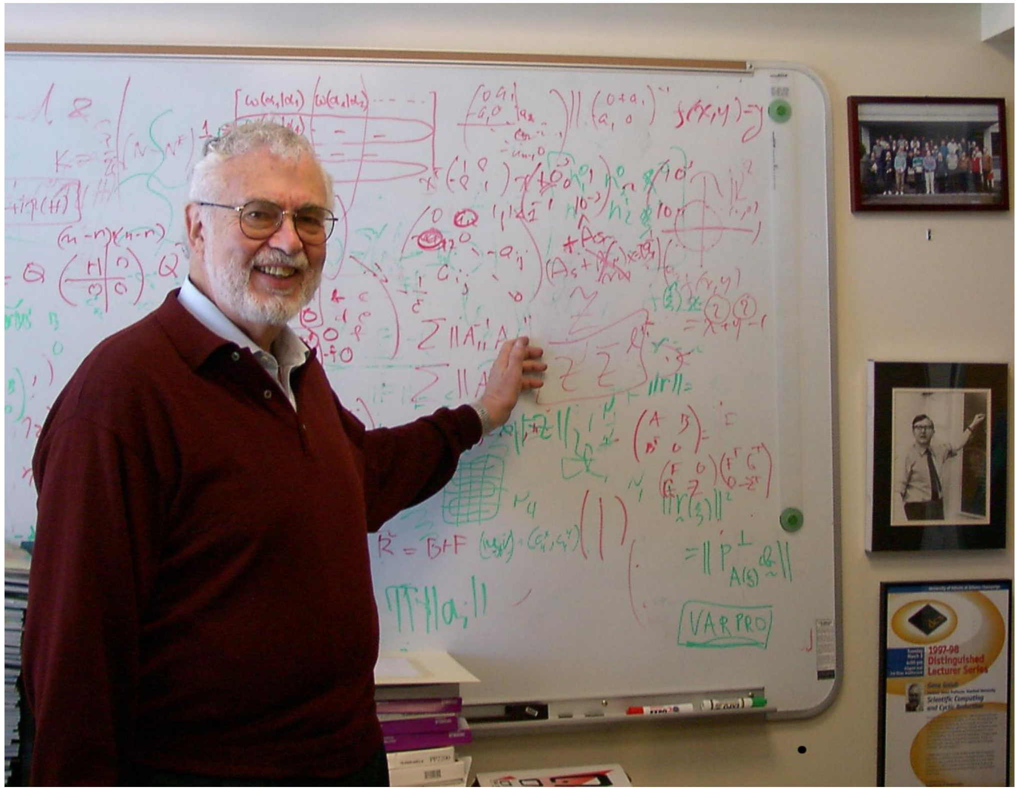
## The Singular Value Decomposition (SVD)

- If $A \in \mathbb{R}^{m\times n}$ (for simplicity: $m \geq n$) then

$$A = U\Sigma V^T$$

- $U \in \mathbb{R}^{m\times m}, \quad V \in \mathbb{R}^{n\times n}$ orthogonal
- $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{m\times n}$ diagonal
- Singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$
- How compute SVD? $\rightarrow$ sophisticated algorithm by Golub-Reinsch

# The Algorithm for Computing the SVD

CHRISTIAN REINSCH GENE GOLUB († 16.11.07)

Singular Value Decomposition and Least Squares Solutions, Numer. Math. 14, 403–420, 1970

# Handbook for Automatic Computation

Edited by

F. L. Bauer · A. S. Householder · F. W. J. Olver
H. Rutishauser † · K. Samelson · E. Stiefel

Volume II

J. H. Wilkinson · C. Reinsch

# Linear Algebra

Chief editor
F. L. Bauer

Springer-Verlag Berlin Heidelberg New York 1971

```
procedure svd (m, n, withu, withv, eps, tol) data: (a) result: (q, u, v);
    value m, n, withu, withv, eps, tol;
    integer m, n;
    Boolean withu, withv;
    real eps, tol;
    array a, q, u, v;
comment Computation of the singular values and complete orthogonal decom-
        position of a real rectangular matrix A,
                A = U diag(q) V^T,    U^T U = V^T V = I,
        where the arrays a[1:m, 1:n], u[1:m, 1:n], v[1:n, 1:n], q[1:n] re-
        present A, U, V, q respectively. The actual parameters corresponding
        to a, u, v may all be identical unless withu = withv = true. In this
        case, the actual parameters corresponding to u and v must differ.
        m ≧ n is assumed;
begin
    integer i, j, k, l, l1;
    real c, f, g, h, s, x, y, z;
    array e[1:n];
    for i := 1 step 1 until m do
    for j := 1 step 1 until n do u[i, j] := a[i, j];
comment Householder's reduction to bidiagonal form;
    g := x := 0;
    for i := 1 step 1 until n do
    begin
        e[i] := g; s := 0; l := i + 1;
        for j := i step 1 until m do s := s + u[j, i]↑2;
        if s < tol then g := 0 else
        begin
            f := u[i, i]; g := if f < 0 then sqrt(s) else -sqrt(s);
            h := f × g - s; u[i, i] := f - g;
            for j := l step 1 until n do
            begin
                s := 0;
                for k := i step 1 until m do s := s + u[k, i] × u[k, j];
                f := s/h;
                for k := i step 1 until m do u[k, j] := u[k, j] + f × u[k, i]
            end j
        end s;
```

## Language Barrier Europe – USA

- Europe: ALCOR-group: exchange of ALGOL programs on punch-cards or paper-tape in the sixties
- USA: Handbook procedures reprogrammed in FORTRAN

  > *The code itself has to be in FORTRAN, which is the language for scientific programming in the United States.*[a]

- The LINPACK project was executed at the end of the seventies at Argonne National Laboratory
- SVD was still a challenge . . .

[a] citation from the preface of the LINPACK users guide

From moler Sun Aug 26 08:22:20 1984
Date: Sun, 26 Aug 84 08:22 PDT
From: Cleve Moler <moler@Navajo> Subject: MATLAB
To: gander@Navajo, higham@Navajo
Status: R

Comments on MATLAB from both Higham and Gander reached me this morning, so I decided to log in on Navajo and use the version of MATLAB here for myself.

Walter is correct. This version of MATLAB cannot find the svd of A = <1 0; 0 1; 1 1>. I suspect it is running into a difficulty where underflow on the VAX with its small floating point exponent is causing loss of information in the implicit shift and hence preventing convergence. svd(100*A) works! This has been corrected in more recent versions.

Nick is also correct about IF's within WHILE's not working correctly. I've know about this for quite a while, but have not yet fixed it.

Instead of flops(2) = 0, use flops = <0 0>.

I didn't know about the unwanted echo into DIARY. Thanks

Do you guys want an up-to-date version of MATLAB on this machine? Other Stanford machines? Who should install it? I don't want to rely on the network to send the complete source code. And I don't know how to send just the corrections that would bring you up to date. (One other thing that doesn't work on your version is handling of overflow. Any overflow puts you in an infinite loop in the error handler.)

So, we should probably send you a new "tar" tape. I'll be glad to do that, if somebody will be responsible for the actual installation. It's not much work.

Reply to na.moler@su-score, since I don't log in here regularly.

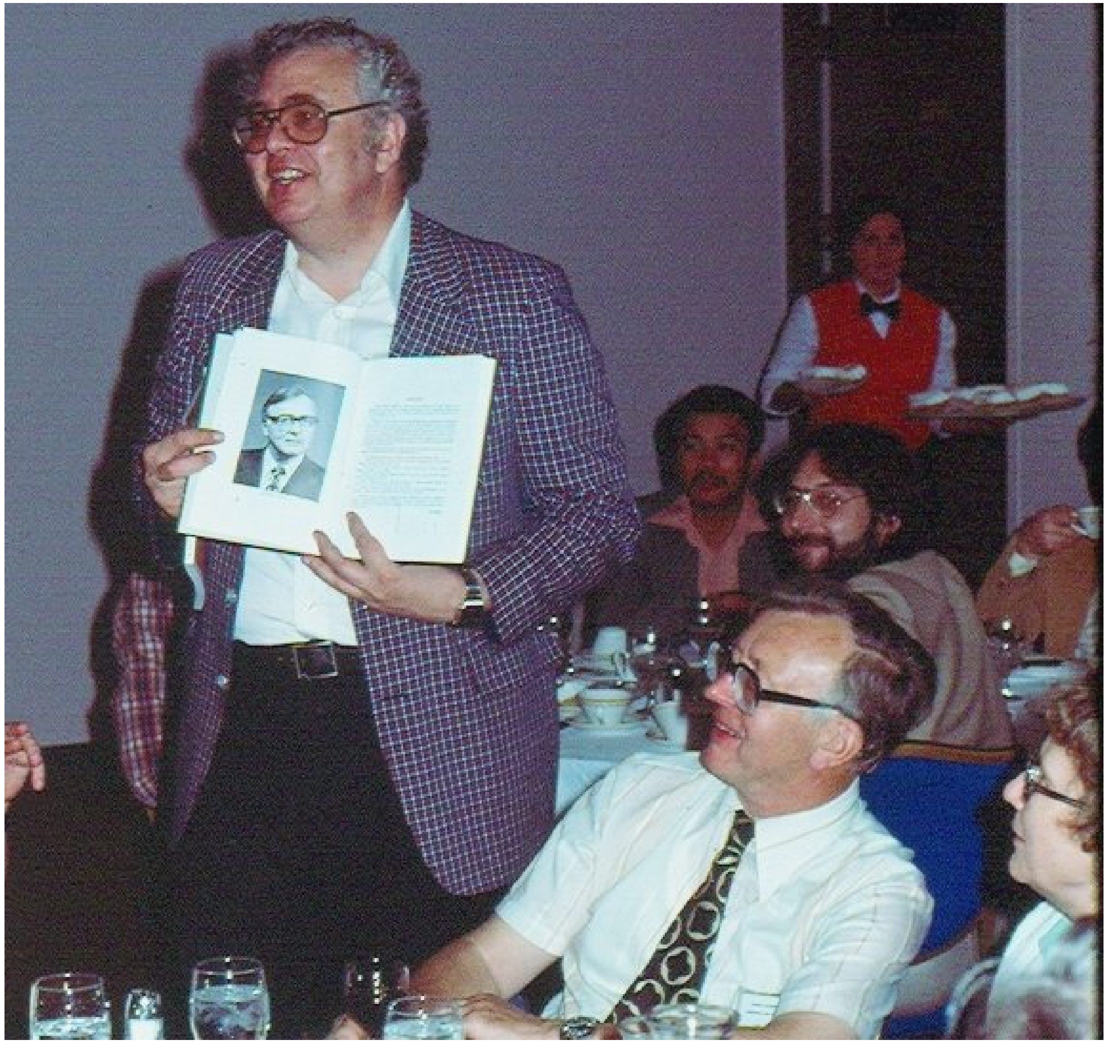
-Cleve


Cleve Moler and Walter Gander in Gene Golub's garden

# 60th Birthday of Jim Wilkinson 1980

Gene presents his new SIAM Journal on Scientific and Statistical Computing. . .

. . . with a dedication to Jim Wilkinson



# Back to SVD

- If $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$ then $r = \text{rank}(A)$
- Partition

$$U = [U_1, U_2] \quad U_1 = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \quad U_2 = [\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m]$$

$$V = [V_1, V_2] \quad V_1 = [\mathbf{v}_1, \ldots, \mathbf{v}_r] \quad V_2 = [\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n]$$

$$\Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$$

- full decomposition

$$A = [U_1, U_2] \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} [V_1, V_2]^T$$

- reduced decomposition (sum of rank-1 matrices)

$$A = U_1 \Sigma_r V_1^T = \mathbf{u}_1 \mathbf{v}_1^T \, \sigma_1 + \mathbf{u}_2 \mathbf{v}_2^T \, \sigma_2 + \cdots + \mathbf{u}_r \mathbf{v}_r^T \, \sigma_r$$

- Fundamental subspaces: SVD provides basis
  - $U_1$ orth. basis for $\mathcal{R}(\mathcal{A})$ (range of $A$)
  $$\mathbf{y} = A\mathbf{x} = U\Sigma V^\top \mathbf{x} = U_1 \Sigma_r \mathbf{z}_r, \quad \mathbf{z} = V^\top \mathbf{x}.$$
  - $V_2$ orth. basis for $\mathcal{N}(\mathcal{A})$ (nullspace)
  $$AV_2 = U\Sigma V^\top V_2 = U\Sigma \binom{0}{I} = U 0_{m\times n} = 0.$$
  - $V_1$ orth. basis for $\mathcal{R}(\mathcal{A}^\top)$
  - $U_2$ orth. basis for $\mathcal{N}(\mathcal{A}^\top)$
- $\mathbb{R}^n = \mathcal{R}(\mathcal{A}^\top) \oplus \mathcal{N}(\mathcal{A}) \qquad \mathbb{R}^m = \mathcal{R}(\mathcal{A}) \oplus \mathcal{N}(\mathcal{A}^\mathcal{T})$
- Projectors

1. $P_{\mathcal{R}(A)} = U_1 U_1^\top$
2. $P_{\mathcal{R}(A^\top)} = V_1 V_1^\top$
3. $P_{\mathcal{N}(A^\top)} = U_2 U_2^\top$
4. $P_{\mathcal{N}(A)} = V_2 V_2^\top$

- Eigenvalues and Singular Values: $\sigma_k(A) = \sqrt{\lambda_k(A^T A)}$
- Matrix Norms

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sigma_1, \quad \|A^{-1}\|_2 = 1/\sigma_n$$

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\text{trace}(A^T A)} = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$$

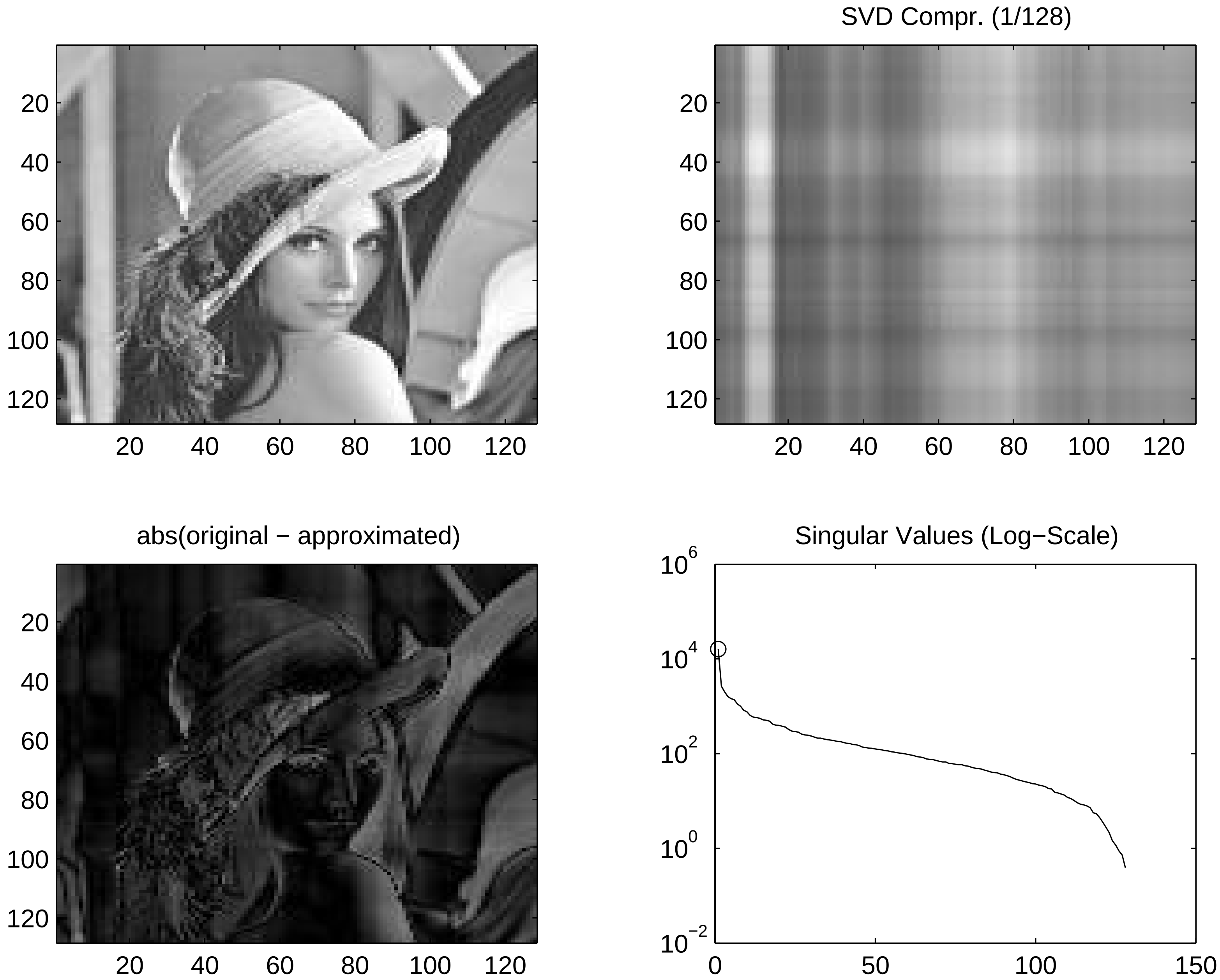
- Condition Number

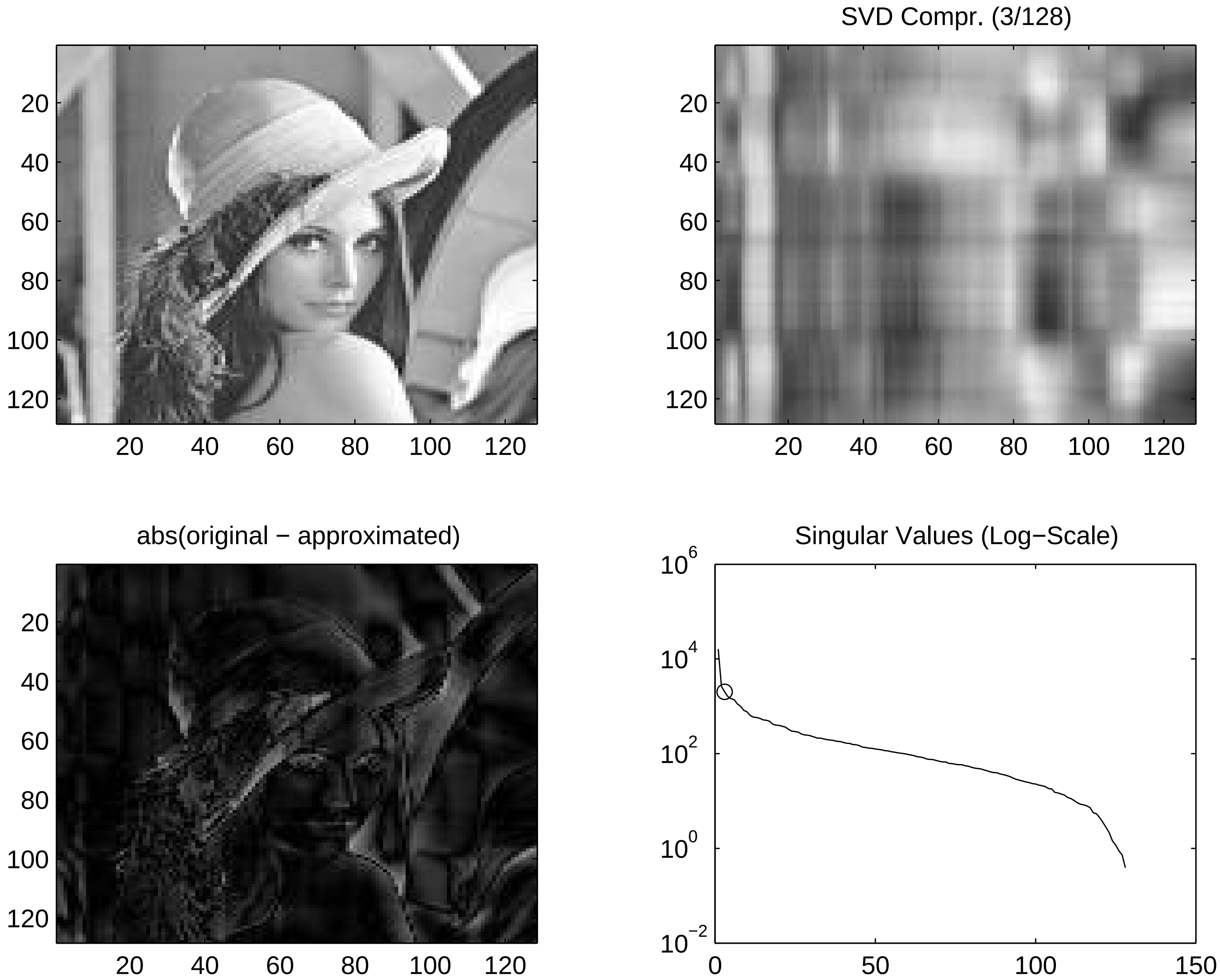
$$\kappa(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$$

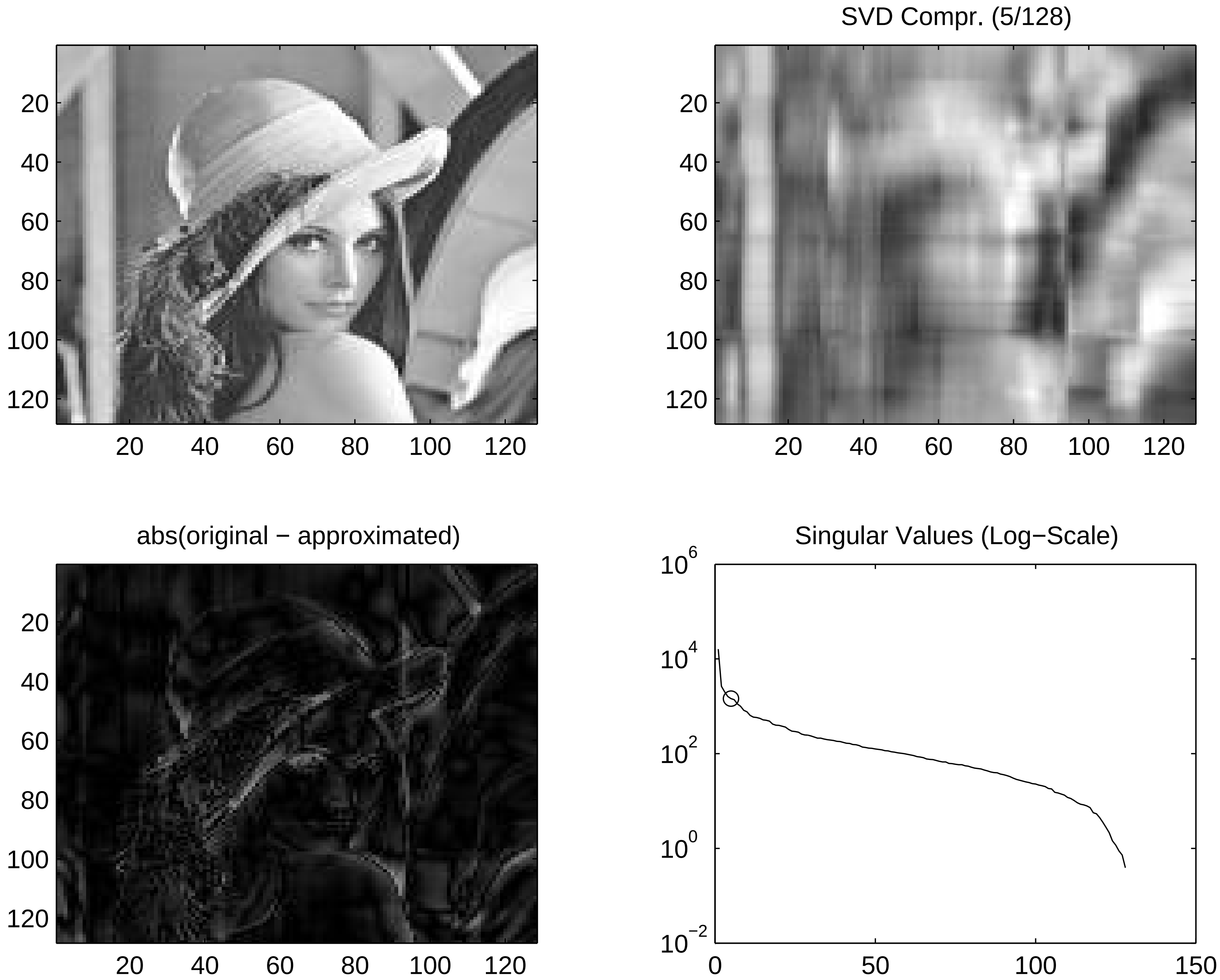
- Lower Rank Approximation $p \leq n$

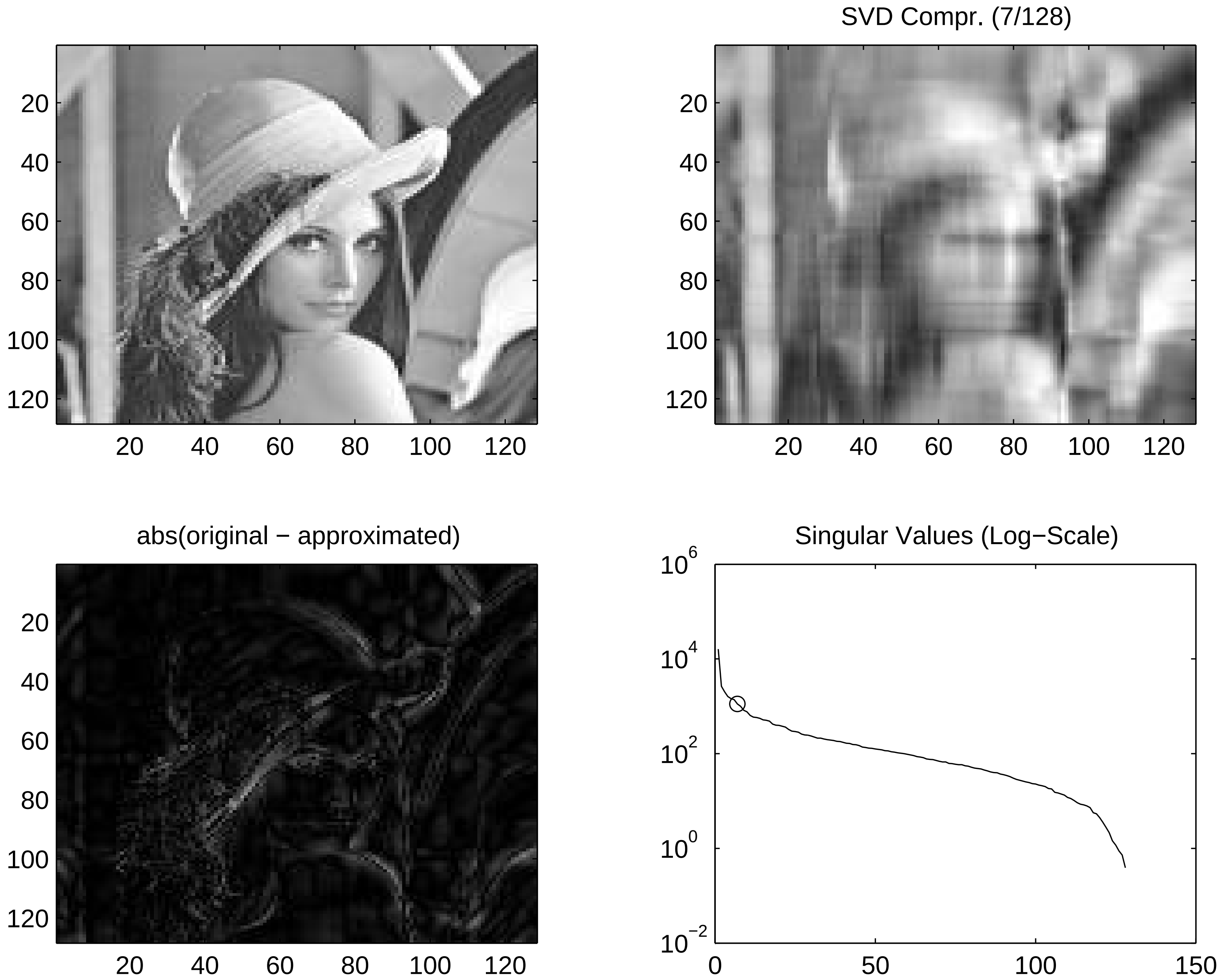
Problem: $\min_{\text{rank}\, X = p} ||A - X||_F$ Solution: $X = \sum_{i=1}^{p} \mathbf{u}_i \mathbf{v}_i^T \sigma_i$

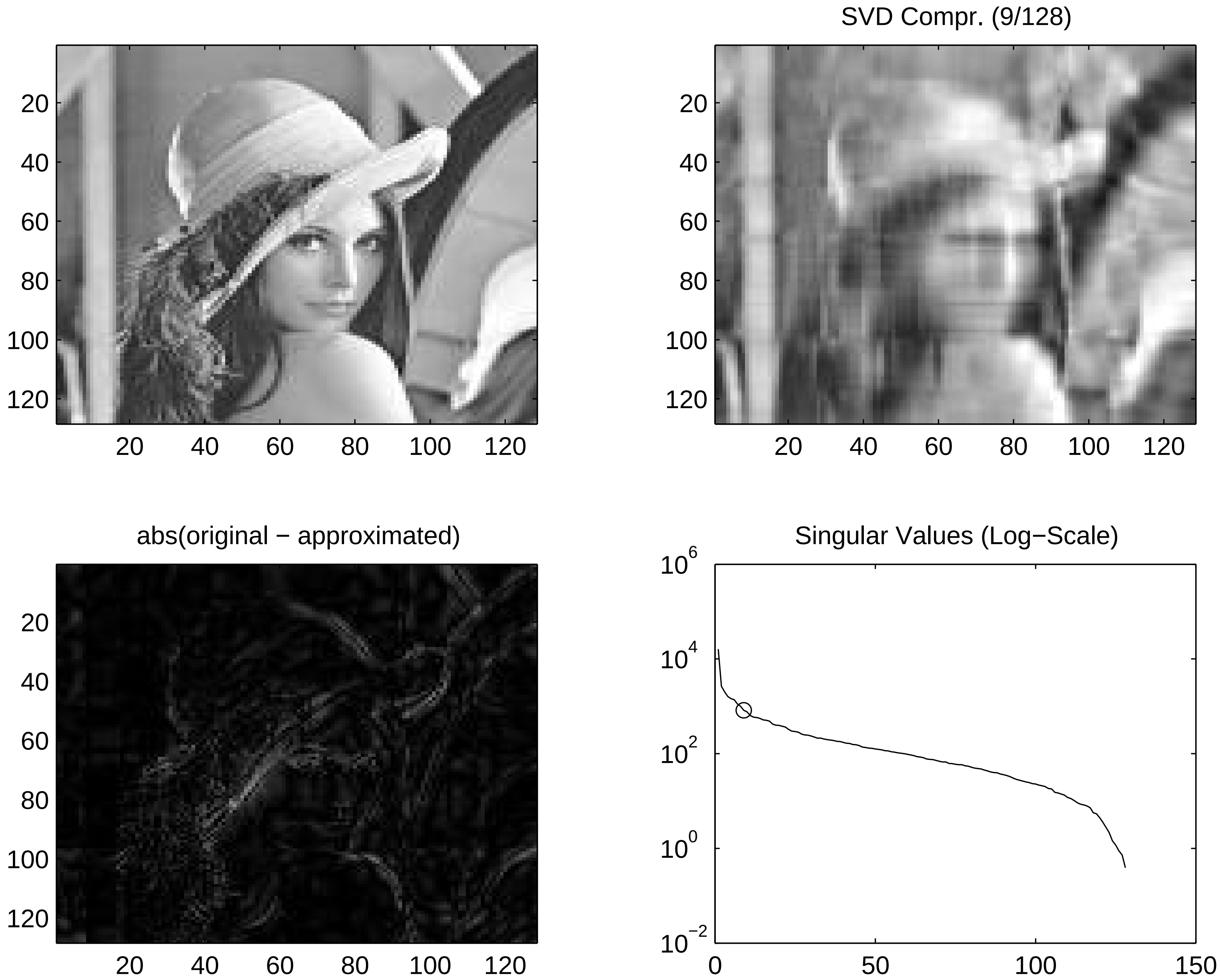
**Lena** $p = 1, 3, 5, 7, 9, 13, 17, 21, 25, 29$

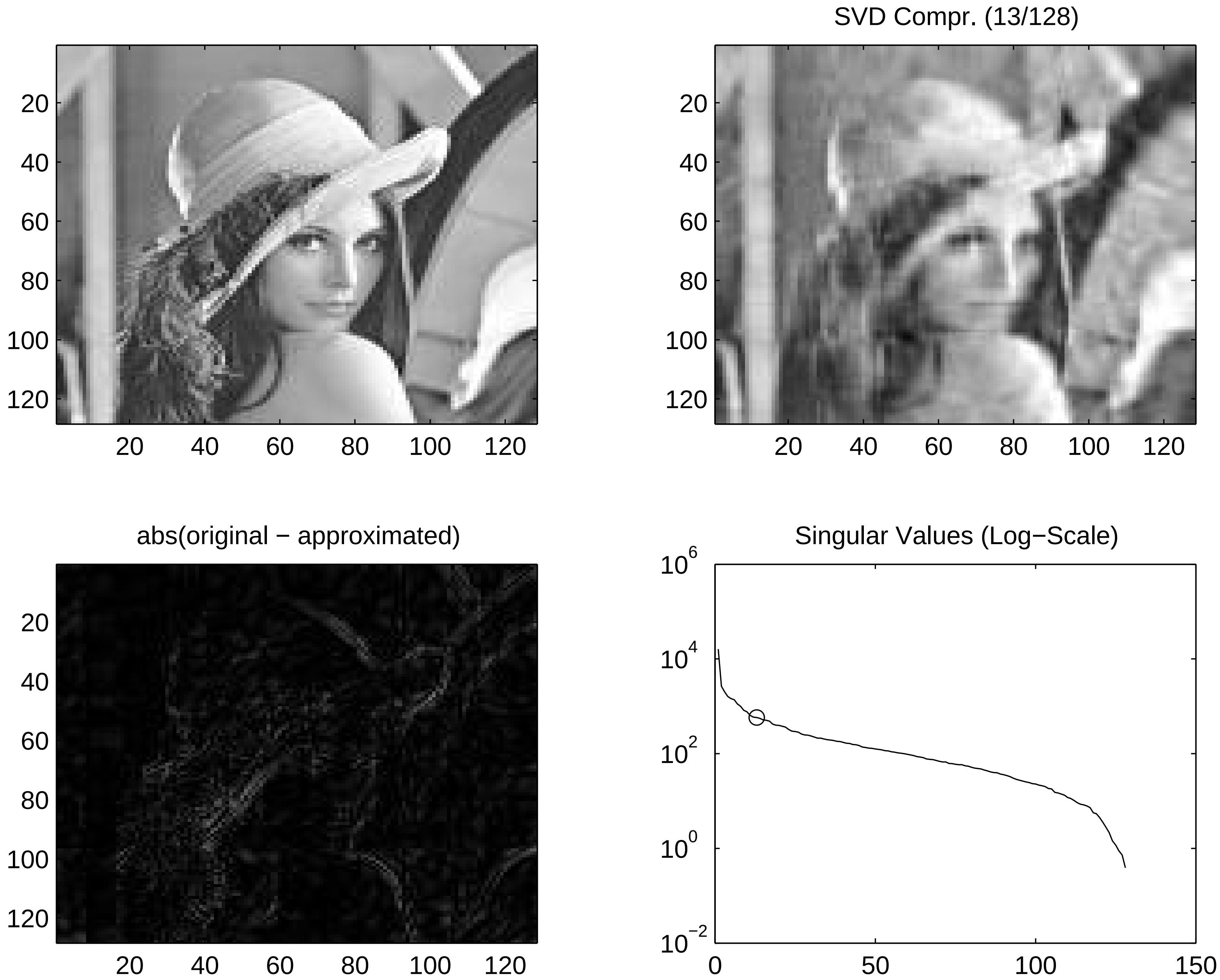

SVD Compr. (1/128)
abs(original − approximated)
Singular Values (Log−Scale)
20
40
60
80
100
120
0
50
100
150
$10^6$
$10^4$
$10^2$
$10^0$
$10^{-2}$


SVD Compr. (3/128)
abs(original − approximated)
Singular Values (Log−Scale)
20
40
60
80
100
120
10^6
10^4
10^2
10^0
10^−2
0
50
100
150

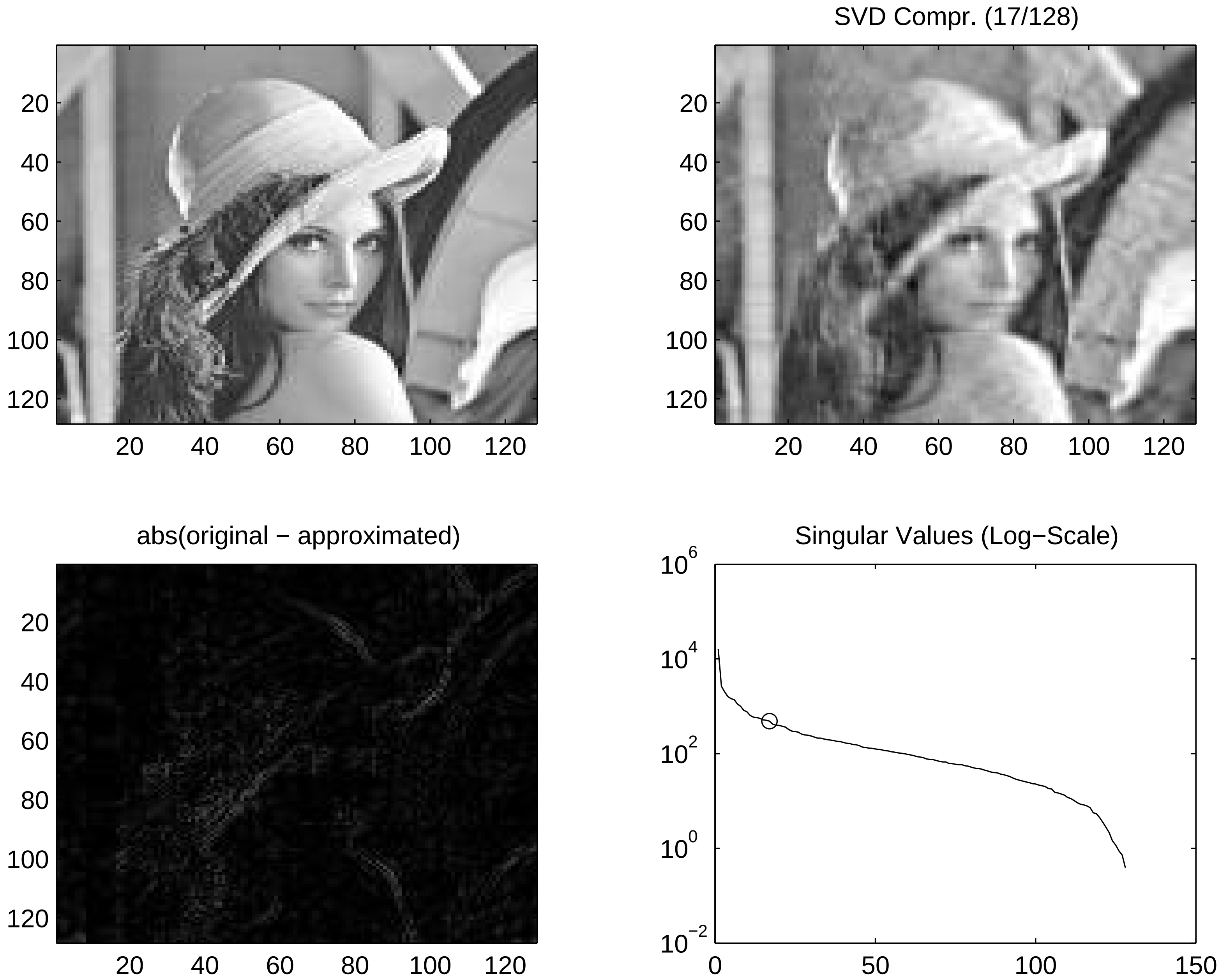

SVD Compr. (5/128)
abs(original − approximated)
Singular Values (Log−Scale)

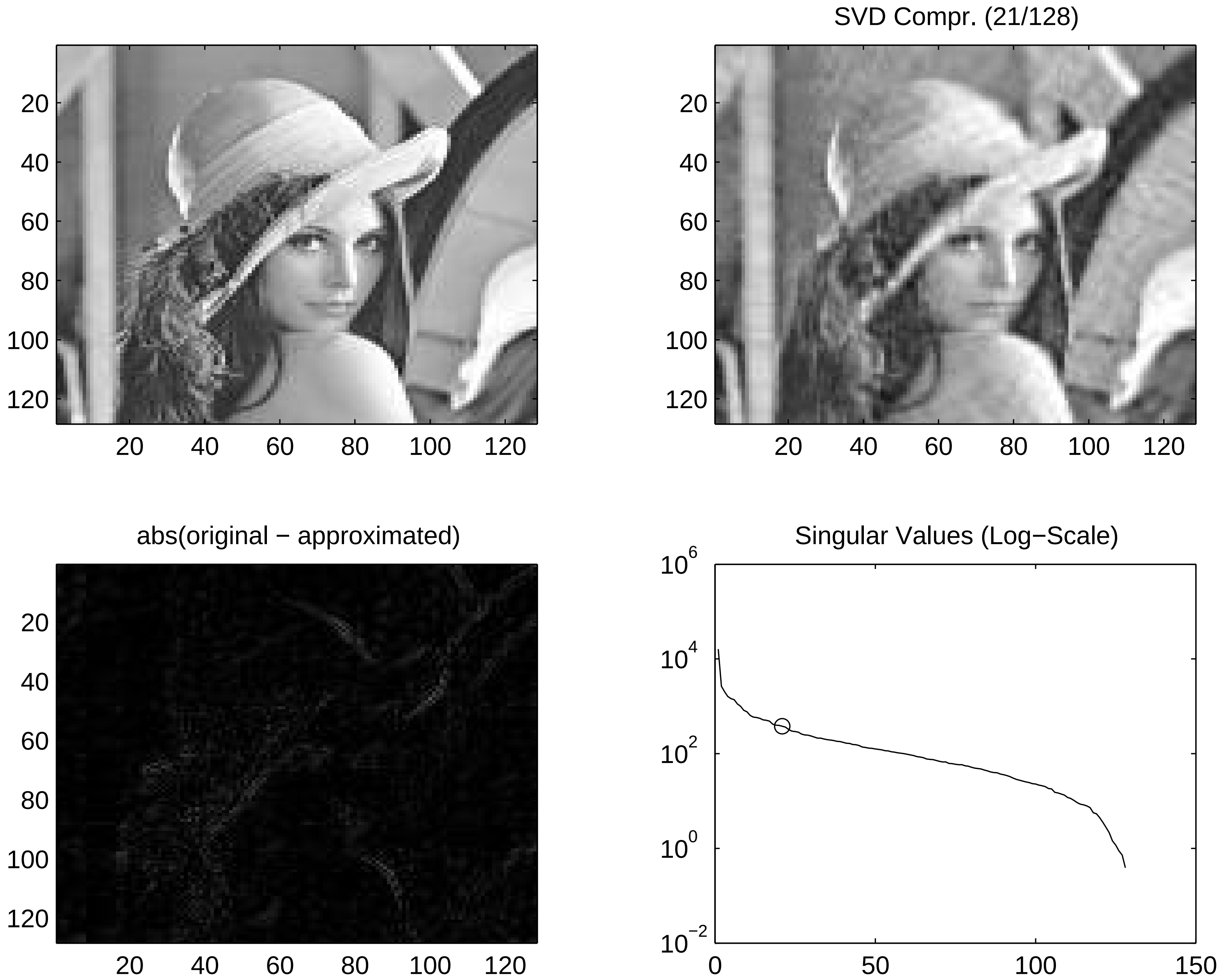

SVD Compr. (7/128)
abs(original − approximated)
Singular Values (Log−Scale)
20 40 60 80 100 120
$10^{6}$ $10^{4}$ $10^{2}$ $10^{0}$ $10^{-2}$
0 50 100 150

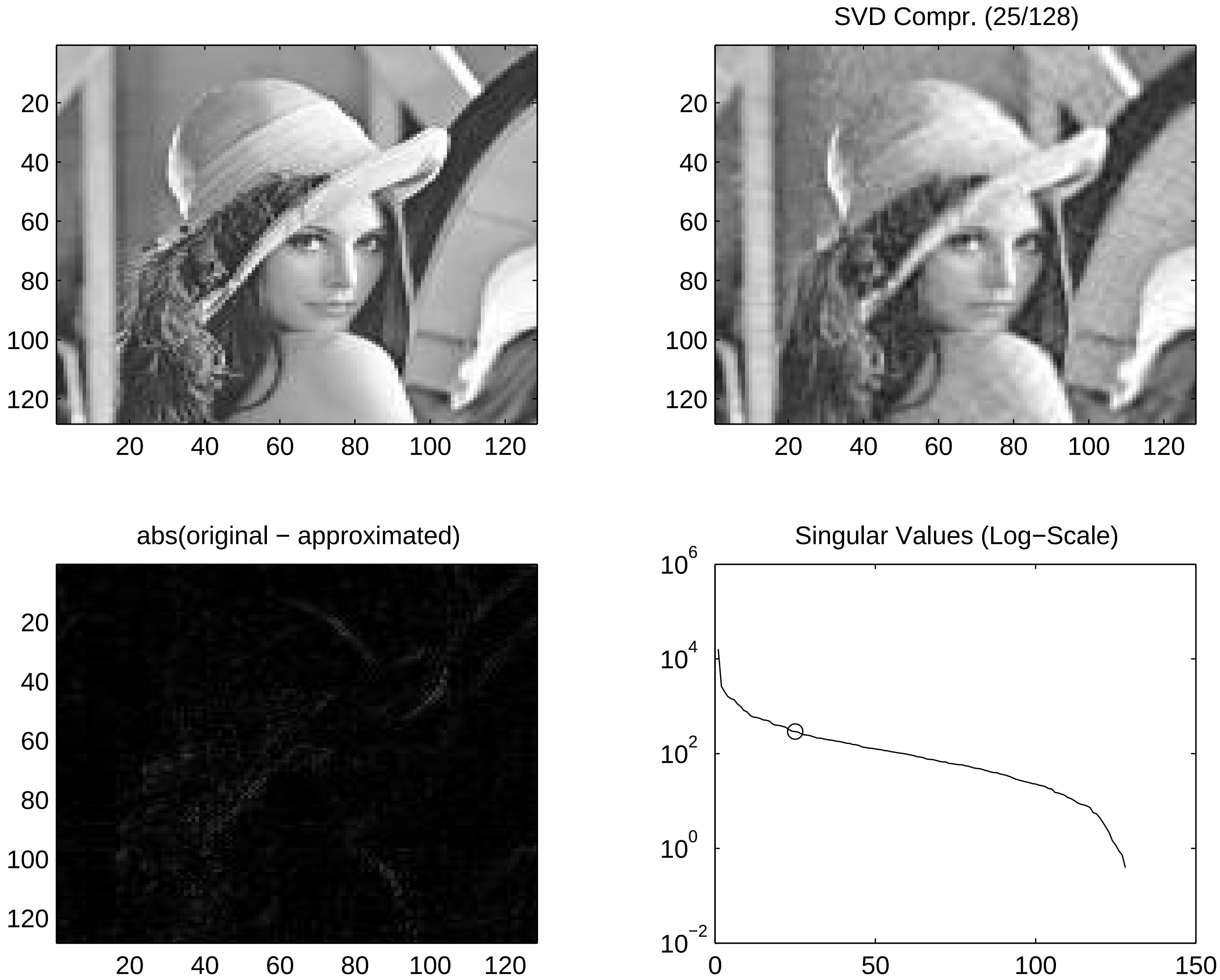

SVD Compr. (9/128)
abs(original − approximated)
Singular Values (Log−Scale)
20 40 60 80 100 120
$10^{6}$ $10^{4}$ $10^{2}$ $10^{0}$ $10^{-2}$
0 50 100 150

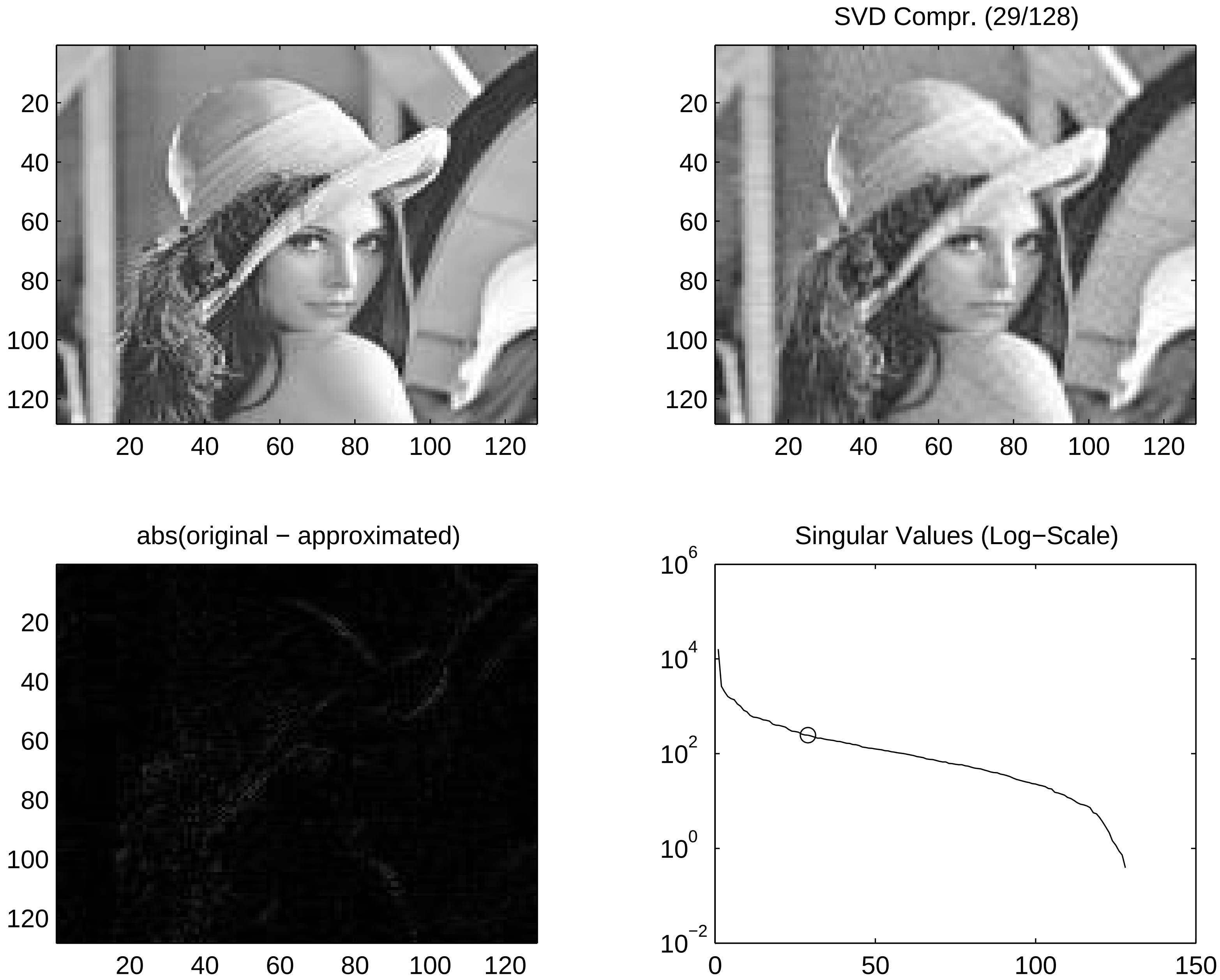


SVD Compr. (13/128)
abs(original − approximated)
Singular Values (Log−Scale)
20
40
60
80
100
120
0
50
100
150
10^6
10^4
10^2
10^0
10^−2


SVD Compr. (17/128)
abs(original − approximated)
Singular Values (Log−Scale)
20
40
60
80
100
120
0
50
100
150
10^6
10^4
10^2
10^0
10^−2


SVD Compr. (21/128)
abs(original − approximated)
Singular Values (Log−Scale)
20
40
60
80
100
120
$10^{6}$
$10^{4}$
$10^{2}$
$10^{0}$
$10^{-2}$
0
50
100
150


SVD Compr. (25/128)
abs(original − approximated)
Singular Values (Log−Scale)
20 40 60 80 100 120
0 50 100 150
$10^{6}$ $10^{4}$ $10^{2}$ $10^{0}$ $10^{-2}$


SVD Compr. (29/128)
abs(original − approximated)
Singular Values (Log−Scale)
20
40
60
80
100
120
$10^6$
$10^4$
$10^2$
$10^0$
$10^{-2}$
0
50
100
150

## General solution of least squares problem $A\mathbf{x} \approx \mathbf{b}$

1. Compute the SVD: `[U,S,V]=svd(A)`.
2. Make a rank decision, i.e. choose $r$ such that $\sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$
3. Partition `V1=V(:,1:r)`, `V2= V(:,r+1:n)`, `Sr=S(1:r,1:r)`, `U1=U(:,1:r)`.
4. The solution with minimal norm is `xm=V1*(Sr\U1'*b)`.
5. The general solution is `x=xm+V2*c` with arbitrary $\mathbf{c} \in \mathbb{R}^{n-r}$.

## Constraint Least Squares

- Let $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$ with $p < n < m$, $\mathrm{rank}(C) = r_c < p$. Consider

$$\|A\mathbf{x} - \mathbf{b}\|_2 \to \min \quad \text{subject to} \quad \|C\mathbf{x} - \mathbf{d}\|_2 \to \min\,.$$

- Compute general solution of $C\mathbf{x} \approx \mathbf{d}$:
  - Compute SVD of $C^\top = TSR^\top$
  - Partition $T = [T_1, T_2]$, $R = [R_1, R_2]$, $S_r = S(1{:}r_c, 1{:}r_c)$
    $T_1 \in \mathbb{R}^{n\times r_c}$, $R_1 \in \mathbb{R}^{p\times r_c}$.

- general solution

$$\mathbf{x} = \mathbf{x}_m + T_2\mathbf{z}, \;\text{with } \mathbf{z} \in \mathbb{R}^{n-r_c} \text{ arbitrary and } \mathbf{x}_m = T_1 S_r^{-1} R_1^\top \mathbf{d}.$$

- Introduce this into $\|A\mathbf{x} - \mathbf{b}\|_2$

## Constraint Least Squares (cont.)

$$\|AT_2\mathbf{z} - (\mathbf{b} - A\mathbf{x}_m)\|_2 \to \min, \quad AT_2 \in \mathbb{R}^{m\times(n-r_c)},$$

- compute the SVD of $AT_2 = U\Sigma V^\top$
- Partition $U = [U_1, U_2]$, $\Sigma_r = \Sigma(1 : r_a, 1 : r_a)$ and $V = [V_1, V_2]$ $(\mathrm{rank}(AT_2) = r_a,\ r_a < n - r_c)$
- general solution (with $\mathbf{z}_m = V_1\Sigma_r^{-1}U_1^\top(\mathbf{b} - A\mathbf{x}_m)$)

$$\mathbf{z} = \mathbf{z}_m + V_2\mathbf{w},\ \ \mathbf{w} \in \mathbb{R}^{n-r_c-r_a} \text{ arbitrary}$$

Thus the solution of

$$\|A\mathbf{x} - \mathbf{b}\|_2 \to \min \quad \text{subject to} \quad \|C\mathbf{x} - \mathbf{d}\|_2 \to \min\,.$$

$$\mathbf{x} = \mathbf{x}_m + T_2V_1\Sigma_r^{-1}U_1^\top(\mathbf{b} - A\mathbf{x}_m) + T_2V_2\mathbf{w}$$

with $\mathbf{x}_m = T_1S_r^{-1}R_1^\top\mathbf{d}$ and $\mathbf{w} \in \mathbb{R}^{n-r_c-r_a}$ arbitrary

# Example $\|A\mathbf{x}-\mathbf{b}\|_2 \to \min$ subject to $\|C\mathbf{x}-\mathbf{d}\|_2 \to \min$

$$A = \begin{pmatrix} 5 & -1 & -1 & 6 & 4 & 0 \\ -3 & 1 & 4 & -7 & -2 & -3 \\ 1 & 3 & -4 & 5 & 4 & 7 \\ 0 & 4 & -1 & 1 & 4 & 5 \\ 4 & 2 & 3 & 1 & 6 & -1 \\ 3 & -3 & -5 & 8 & 0 & 2 \\ 0 & -1 & -4 & 4 & -1 & 3 \\ -5 & 4 & -3 & -2 & -1 & 7 \\ 3 & 4 & -3 & 6 & 7 & 7 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} -4 \\ 1 \\ -2 \\ 3 \\ 3 \\ 0 \\ -1 \\ 3 \\ 1 \end{pmatrix}$$

$A \in \mathbb{R}^{9\times 6}$, $\operatorname{rank}(A) = 3$

$C \in \mathbb{R}^{3\times 6}$, $\operatorname{rank}(C) = 2$

$\operatorname{rank}\binom{A}{C} = 5$

$\mathbf{b} \notin \mathcal{R}(A)$

$\mathbf{d} \notin \mathcal{R}(C)$

$$C = \begin{pmatrix} 1 & 3 & -2 & 3 & 8 & 0 \\ -3 & 0 & 0 & 1 & 9 & 4 \\ -2 & 3 & -2 & 4 & 17 & 4 \end{pmatrix} \quad \mathbf{d} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$$

Solution: $$\mathbf{x} = \begin{pmatrix} 0.1073 \\ 0.2230 \\ 0.2695 \\ -0.1950 \\ -0.0815 \\ 0.3126 \end{pmatrix} + \lambda \begin{pmatrix} 0.4656 \\ -0.2993 \\ -0.4989 \\ -0.6319 \\ 0.1663 \\ 0.1330 \end{pmatrix}.$$

# Quadratically Constrained Problems

$$\max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sigma_1, \qquad \min_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sigma_n$$

Thus the problem

$$\|A\mathbf{x}\|_2 = \min, \quad \text{subject to } \|\mathbf{x}\|_2 = 1$$

has the solution

$$\mathbf{x} = \mathbf{v}_n \quad \text{with} \quad \min_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \|A\mathbf{v}_n\|_2 = \sigma_n$$

where $A = U\Sigma V^T$ (Singular Value Decomposition)

$V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ $\qquad \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$

## Example: Fitting Ellipses

- Quadratic Form for Conic Section $\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0, \quad A = A^T$
- unknown coefficients $\mathbf{u} = [a_{11}, a_{12}, a_{22}, b_1, b_2\ , c]$
- normalize the coefficients by $\|\mathbf{u}\|_2 = 1$
- Inserting measured points $\mathbf{x}_i^T A\mathbf{x}_i + \mathbf{b}^T\mathbf{x}_i + c \approx 0$

  $$||B\mathbf{u}||_2 = \min \quad \text{subject to } ||\mathbf{u}||_2 = 1$$

  quadratically constrained problem $\Longrightarrow$ solution by SVD
- Need to compute position/orientation and geometric parameters from $\mathbf{u}$ with principal axes transformation

demo

```
function [z,a,b,alpha]=algellipse(X);
%  [z,a,b]=algellipse(X)
% fits an ellipse by minimizing the "algebraic distance"
% to given points  Pi=[X(i,1),X(i,2)]
% in the least squares sense   x'A x * bb'x + c=0
 [U S V]= svd([X(:,1).^2  X(:,1).*X(:,2) X(:,2).^2 ...
              X(:,1) X(:,2) ones(size(X(:,1)))]);
 u=V(:,6);
 A=[u(1) u(2)/2; u(2)/2 u(3)];
 bb =[u(4); u(5)]; c=u(6);
 [Q D]=eig(A);                 % principle axis transformation
 alpha=atan2(Q(2,1),Q(1,1));
 bs=Q'*bb;
 zs=-(2*D)\bs;  z=Q*zs;
 h=-bs'*zs/2-c;
 a=sqrt(h/D(1,1));
 b=sqrt(h/D(2,2));
```

## Are Points Collinear on a Hyperplane?

- Given $m$ points in $\mathbb{R}^n$: $X^T = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]$
- Compute barycentric coordinates

$$\hat{X}^T = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_m] \quad \text{with } \hat{\mathbf{x}}_i = \mathbf{x}_i - \frac{1}{m}\sum_{i=1}^{m} \mathbf{x}_i$$

- Compute SVD of $\hat{X} = U\Sigma V^T$
- for $s = 1, 2, \ldots, n-1$

$$V_1 = V(:, 1{:}s)$$

| hyperplane | Collinearity Measure: Residual Sum of Squares |
|---|---|
| $\mathbf{y} = \mathbf{p} + V_1\mathbf{t}$ | $\sigma_{s+1}^2 + \sigma_{s+2}^2 + \cdots + \sigma_n^2$ |

```
function [V,p]= hyper(Q);
% HYPER  Fits a hyperplane of dimension s <n to given points Q(i,:)
%        The hyperplane is X=p + V(:,1:s)*tau (Parameter Form) or
%        is defined as solution of the linear
%        equations  V(:,s+1:n)'*(y - p)=0 (Normal form)
m=max(size(Q));
p=sum(Q)'/m;
Qt=Q-ones(size(Q))*diag(p);
[U,S,V]=svd(Qt,0);
```