

3D Computer Vision - WT 2005

Singular Value Decomposition

Martin Groher

CAMP - Chair for Computer Aided Medical Procedures

November 16, 2005



- 1 Singular Value Decomposition (SVD)
- 2 Optimization
- 3 SVD Computation
- 4 Clarifier: Multivariate Calculus for Least Squares
- 5 Summary and References

Singular Value Decomposition (SVD)

- For any given matrix $A \in \mathbf{R}^{m \times n}$ there exists a decomposition

$$A = UDV^T$$

such that

- U is an $m \times n$ matrix with orthogonal columns
- D is a $n \times n$ diagonal matrix with non-negative entries
- V^T is an $n \times n$ orthogonal matrix

SVD - Visualized

$$\begin{bmatrix} A \end{bmatrix} = \begin{bmatrix} U \end{bmatrix} \begin{bmatrix} D \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix}$$

- A is an $m \times n$ matrix
- U is an $m \times m$ matrix with orthogonal columns
- D is a $n \times n$ diagonal matrix with non-negative entries
- V^T is an $n \times n$ orthogonal matrix

Singular Value Decomposition (SVD)

- The diagonal values of D are called **Singular Values** of A
- The column vectors of U are the **Left Singular Vectors** of A
- The column vectors of V are the **Right Singular Vectors** of A .

Properties

- The SVD can be performed s.t. the diagonal values of D are descending i.e. $d_1 \geq d_2 \geq \dots \geq d_n \geq 0$.
- We will assume that the SVD is always performed in that way
- The diagonal values of D are the square roots of the Eigenvalues of $A^T A$ and AA^T
(Hence the non-negativity of the elements of D)

Some More Properties

- It holds for the left singular vectors u_i :

$$A^T A u_i = d_i^2 u_i$$

- It holds for the right singular vectors v_i :

$$A A^T v_i = d_i^2 v_i$$

- The left singular vectors u_i are eigenvectors of $A^T A$
- The right singular vectors v_i are eigenvectors of $A A^T$

Even More Properties

- SVD explicitly constructs orthonormal bases for the null-space and the range of a matrix
- columns of U corresponding to non-zero elements of D span the range
- columns of V corresponding to zero elements of D span the null-space

Properties Galore

- SVD allows a rank decision:

$rank(A)$ is the largest r s.t. $d_r > 0$

- there are $m - r$ left singular vectors corresponding to the singular value 0.
- there are $n - r$ right singular vectors corresponding to the singular value 0.

Linear Optimization

SVD can be used for linear optimization by using the following property

- Let v_n be the right singular vector corresponding to d_n (the smallest element of D)
- The product Ax with $\|x\|_2 = 1$ has the minimal value for $x = v_n$.

Minimization by SVD

The minimizing property of the last right singular vector v_n can be used to solve the following minimization task

- Given the linear function $f = Ax, f : \mathbf{R}^n \rightarrow \mathbf{R}^m$ to be minimized (in most applications $m \gg n$)
- With the constraint that the solution x is not trivial ($x \neq 0$)
(We will assume that $\|x\|_2 = 1$)

Minimization by SVD II

- The minimization problem is thus

$$\text{minimize } Ax \quad \text{s.t. } \|x\|_2 = 1$$

- It can be shown that the solution is the right singular vector $x = v_n$ corresponding to the smallest singular value d_n .

Proof

- Problem: minimize $\|Ax\|_2$ subject to $\|x\|_2 = 1$
- Because of the orthogonality of U and V we have
 - $\|Ax\|_2 = \|UDV^T x\|_2 = \|DV^T x\|_2$
 - $\|x\|_2 = \|V^T x\|_2$
- Hence have to minimize $\|DV^T x\|_2$ subject to $\|V^T x\|_2 = 1$
- With $V^T x = y$ we have: minimize $\|Dy\|_2$ subject to $\|y\|_2 = 1$
- Since D diagonal with descending entries we get
 $y = (0, 0, \dots, 0, 1)^T$
- Since $V^T x = y \Leftrightarrow x = Vy$ we get $x = v_n$

What's next...?

So in order to solve a linear minimization problem by SVD we have to do two things

- 1 State it in the form

$$\text{minimize } Ax \quad \text{s.t.} \quad \|x\|_2 = 1$$

- 2 Compute the SVD $A = UDV^T$ and take the last right singular vector v_n as the solution

Notes

- Minimization by SVD is extremely simple. The tricky part is to state the problem the right way.
- The error minimized by the SVD is called **Algebraic Error** or **Algebraic Distance**
- A drawback of the algebraic error is that it is geometrically meaningless, so that minimizing it can lead to completely meaningless results

SVD Black Box

- Only thing left is to show how an SVD of a matrix A can be computed
- But we won't do that.
- We'll use SVD as a black box
- Computation involves QR procedure and Householder reduction
- Original algorithm by Golub and Reinsch

Computation Properties

- Algorithm is extremely stable
- Computation time for SVD of an $m \times n$ A :
 - Computation of U , V and D : $4m^2n + 8mn^2 + 9n^3$
 - Computation of V and D : $4mn^2 + 8n^3$
- Keep in mind that in most cases $m \gg n$

Implementation in Matlab

- Matlab has the command $[U,S,V] = \text{SVD}(X)$
- Attention: The command returns V and not V^T
- Hence it holds that $X = U*S*V'$

Multivariate Calculus for Least Squares

- Proof that $\|Ax\|$ is minimized w.s.t. $\|x\| = 1$ may not be intuitively plausible
- We will apply multivariate calculus to make it more plausible

What we have learned



$$f(x_1, \dots, x_n) : \mathbb{R}^n \rightarrow \mathbb{R} \Rightarrow \text{grad}(f(x_1, \dots, x_n)) = \begin{pmatrix} \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \end{pmatrix}$$



$$g(x_1, \dots, x_n) : \mathbb{R}^n \rightarrow \mathbb{R}^m \Rightarrow J_g = \begin{pmatrix} \text{grad}(g_1(x_1, \dots, x_n))^T \\ \text{grad}(g_2(x_1, \dots, x_n))^T \\ \vdots \\ \text{grad}(g_n(x_1, \dots, x_n))^T \end{pmatrix}$$

Building derivatives without losing vector notation

- Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$; $\mathbf{h} \in \mathbb{R}^n$; $\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$
- for instance: $f(\mathbf{h}) = \|\mathbf{h}\|^2 = \mathbf{h}^T \mathbf{h} = h_1^2 + \dots + h_n^2$.
- $\Rightarrow f'(\mathbf{h}) = \begin{pmatrix} \frac{\partial \mathbf{h}^T \mathbf{h}}{\partial h_1} \\ \vdots \\ \frac{\partial \mathbf{h}^T \mathbf{h}}{\partial h_n} \end{pmatrix} = \begin{pmatrix} 2h_1 \\ \vdots \\ 2h_n \end{pmatrix} = 2\mathbf{h}$
-

$$\begin{aligned} \mathbf{h}^T \mathbf{h} &\cong x^2 \\ f'(\mathbf{h}) = 2\mathbf{h} &\cong f'(x) = 2x \end{aligned}$$

Building derivatives without losing vector notation

- for instance:

$$f(\mathbf{h}) = \|\mathbf{A}\mathbf{h}\|^2 = \mathbf{h}^T \underbrace{\mathbf{A}^T \mathbf{A}}_{\mathbf{C}} \mathbf{h}, \quad \mathbf{h} \in \mathbb{R}^2, \quad \mathbf{A} \in \mathbb{R}^{2 \times 2}.$$

- $f(\mathbf{h}) = (h_1, h_2) \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$

- $\Rightarrow f'(\mathbf{h}) = h_1^2 c_{11} + 2h_1 h_2 c_{12} + h_2^2 c_{22}$
since $c_{12} = c_{21}$, $\mathbf{A}^T \mathbf{A}$ symmetric!

- $\Rightarrow f'(\mathbf{h}) = \begin{pmatrix} 2h_1 c_{11} + 2h_2 c_{12} \\ 2h_1 c_{22} + 2h_2 c_{22} \end{pmatrix} = 2\mathbf{A}^T \mathbf{A} \mathbf{h}.$

Using multivariate calculus for linear optimization: SVD

- Minimize $\|\mathbf{Ax}\|$ subject to $\|\mathbf{x}\| = 1$
- \Leftrightarrow Minimize $\|\mathbf{Ax}\|^2$ subject to $\|\mathbf{x}\|^2 = 1$
- \Leftrightarrow Minimize $\mathbf{x}^T \mathbf{A}^T \mathbf{Ax}$ subject to $\mathbf{x}^T \mathbf{x} = 1$
- \Leftrightarrow Minimize $\underbrace{\mathbf{x}^T \mathbf{A}^T \mathbf{Ax}}_{f(\mathbf{x})}$ subject to $-\underbrace{\mathbf{x}^T \mathbf{x} + 1}_{g(\mathbf{x})} = 0$
- Minimize both $f(\mathbf{x})$ and $g(\mathbf{x})$ using Lagrange-multiplier λ
 - $\Leftrightarrow \min(f(\mathbf{x}) + \lambda g(\mathbf{s}))$
 - $\Leftrightarrow \min(\mathbf{x}^T \mathbf{A}^T \mathbf{Ax} + \lambda(1 - \mathbf{x}^T \mathbf{x}))$

Using multivariate calculus for linear optimization: SVD

- How to Minimize? Build 1st derivative and set to zero:
- $\Leftrightarrow 2\mathbf{A}^T \mathbf{A} \mathbf{x} - 2\lambda \mathbf{x} = 0$
- $\Leftrightarrow \mathbf{A}^T \mathbf{A} \mathbf{x} = \lambda \mathbf{x}$
- \Leftrightarrow Find the eigenvector associated to the smallest eigenvalue of $\mathbf{A}^T \mathbf{A}$
- \Leftrightarrow Find the singular vector associated to the smallest value of \mathbf{A} (**SVD!!**)

Summary

- Properties of the SVD
- Linear minimization using SVD
- Plausible explanation of Least Squares with SVD

References

- Hartley and Zisserman. *Computer Vision*
- Walter Gander. *Fitting Data by Least Squares - Algorithms and Examples*
- Gander and Hrebicek. *Solving Problems in Scientific Computing using Maple and Matlab*
- Vachenaer. *Höhere Mathematik 1*
- Press et. al. *Numerical Recipes in C*
- Golub and van Loan. *Matrix Computations*