# Least-squares Solution of Homogeneous Equations

supportive text for teaching purposes

Revision: 1.2, dated: December 15, 2005

#### Tomáš Svoboda

Czech Technical University, Faculty of Electrical Engineering Center for Machine Perception, Prague, Czech Republic

svoboda@cmp.felk.cvut.cz

http://cmp.felk.cvut.cz/~svoboda

#### Introduction

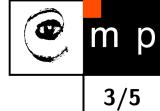
We want to find a  $n \times 1$  vector **h** satisfying

$$\mathbf{Ah} = \mathbf{0}$$
,

where A is  $m \times n$  matrix, and 0 is  $n \times 1$  zero vector. Assume  $m \ge n$ , and  $\operatorname{rank}(A) = n$ . We are obviously not interested in the trivial solution  $\mathbf{h} = \mathbf{0}$  hence, we add the constraint

$$\|\mathbf{h}\| = 1$$
.

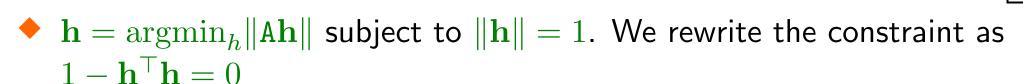
Constrained least–squares minimization: Find  ${\bf h}$  that minimizes  $\|{\bf A}{\bf h}\|$  subject to  $\|{\bf h}\|=1$ .



•  $\mathbf{h} = \operatorname{argmin}_h \|\mathbf{A}\mathbf{h}\|$  subject to  $\|\mathbf{h}\| = 1$ . We rewrite the constraint as  $1 - \mathbf{h}^{\top}\mathbf{h} = 0$ 

- $\mathbf{h} = \operatorname{argmin}_h \|\mathbf{A}\mathbf{h}\|$  subject to  $\|\mathbf{h}\| = 1$ . We rewrite the constraint as  $1 \mathbf{h}^{\top}\mathbf{h} = 0$
- To find an extreme (the sought h) we must solve  $\frac{\partial}{\partial \mathbf{h}} (\mathbf{h}^{\top} \mathbf{A}^{\top} \mathbf{A} \mathbf{h} + \lambda (1 \mathbf{h}^{\top} \mathbf{h})) = 0$ .

- $\mathbf{h} = \operatorname{argmin}_h \|\mathbf{A}\mathbf{h}\|$  subject to  $\|\mathbf{h}\| = 1$ . We rewrite the constraint as  $1 \mathbf{h}^{\top}\mathbf{h} = 0$
- To find an extreme (the sought h) we must solve  $\frac{\partial}{\partial \mathbf{h}} \left( \mathbf{h}^{\top} \mathbf{A}^{\top} \mathbf{A} \mathbf{h} + \lambda (1 \mathbf{h}^{\top} \mathbf{h}) \right) = 0$ .
- We derive:  $2\mathbf{A}^{\top}\mathbf{A}\mathbf{h} 2\lambda\mathbf{h} = 0$ .



- To find an extreme (the sought  $\mathbf{h}$ ) we must solve  $\frac{\partial}{\partial \mathbf{h}} \left( \mathbf{h}^{\top} \mathbf{A}^{\top} \mathbf{A} \mathbf{h} + \lambda (1 \mathbf{h}^{\top} \mathbf{h}) \right) = 0$ .
- We derive:  $2\mathbf{A}^{\top}\mathbf{A}\mathbf{h} 2\lambda\mathbf{h} = 0$ .
- After some manipulation we end up with:  $(\mathbf{A}^{\top}\mathbf{A} \lambda\mathbf{E})\mathbf{h} = 0$  which is the characteristic equation. Hence, we know that  $\mathbf{h}$  is an eigenvector of  $(\mathbf{A}^{\top}\mathbf{A})$  and  $\lambda$  is an eigenvalue.





- $\mathbf{h} = \operatorname{argmin}_h \|\mathbf{A}\mathbf{h}\|$  subject to  $\|\mathbf{h}\| = 1$ . We rewrite the constraint as  $1 \mathbf{h}^{\top}\mathbf{h} = 0$
- To find an extreme (the sought  $\mathbf{h}$ ) we must solve  $\frac{\partial}{\partial \mathbf{h}} \left( \mathbf{h}^{\top} \mathbf{A}^{\top} \mathbf{A} \mathbf{h} + \lambda (1 \mathbf{h}^{\top} \mathbf{h}) \right) = 0$ .
- We derive:  $2\mathbf{A}^{\top}\mathbf{A}\mathbf{h} 2\lambda\mathbf{h} = 0$ .
- After some manipulation we end up with:  $(\mathbf{A}^{\top}\mathbf{A} \lambda\mathbf{E})\mathbf{h} = 0$  which is the characteristic equation. Hence, we know that  $\mathbf{h}$  is an eigenvector of  $(\mathbf{A}^{\top}\mathbf{A})$  and  $\lambda$  is an eigenvalue.
- The least-squares error is  $e = \mathbf{h}^{\top} \mathbf{A}^{\top} \mathbf{A} \mathbf{h} = \mathbf{h}^{\top} \lambda \mathbf{h}$ .

- $\mathbf{h} = \operatorname{argmin}_h \|\mathbf{A}\mathbf{h}\|$  subject to  $\|\mathbf{h}\| = 1$ . We rewrite the constraint as  $1 \mathbf{h}^{\top}\mathbf{h} = 0$
- To find an extreme (the sought  $\mathbf{h}$ ) we must solve  $\frac{\partial}{\partial \mathbf{h}} \left( \mathbf{h}^{\top} \mathbf{A}^{\top} \mathbf{A} \mathbf{h} + \lambda (1 \mathbf{h}^{\top} \mathbf{h}) \right) = 0$ .
- We derive:  $2\mathbf{A}^{\top}\mathbf{A}\mathbf{h} 2\lambda\mathbf{h} = 0$ .
- After some manipulation we end up with:  $(\mathbf{A}^{\top}\mathbf{A} \lambda\mathbf{E})\mathbf{h} = 0$  which is the characteristic equation. Hence, we know that  $\mathbf{h}$  is an eigenvector of  $(\mathbf{A}^{\top}\mathbf{A})$  and  $\lambda$  is an eigenvalue.
- The least-squares error is  $e = \mathbf{h}^{\top} \mathbf{A}^{\top} \mathbf{A} \mathbf{h} = \mathbf{h}^{\top} \lambda \mathbf{h}$ .
- The error will be minimal for  $\lambda = \min_i \lambda_i$  and the sought solution is then the eigenvector of the matrix  $(A^T A)$  corresponding to the smallest eigenvalue.

#### Derivation II — SVD



4/5

• Let  $A = USV^{\top}$ , where U is  $m \times n$  orthonormal, S is  $n \times n$  diagonal with descending order, and  $V^{\top}$  is  $n \times n$  also orthonormal.



4/5

- Let  $A = USV^{\top}$ , where U is  $m \times n$  orthonormal, S is  $n \times n$  diagonal with descending order, and  $V^{\top}$  is  $n \times n$  also orthonormal.
- From orthonormality of U, V follows that  $\|\mathbf{USV}^{\top}\mathbf{h}\| = \|\mathbf{SV}^{\top}\mathbf{h}\|$  and  $\|\mathbf{V}^{\top}\mathbf{h}\| = \|\mathbf{h}\|$ .

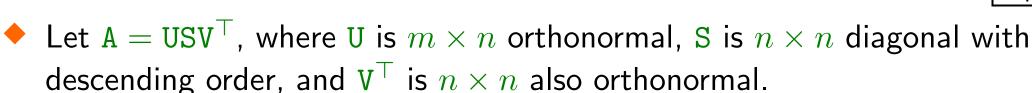
#### **Derivation II — SVD**

- Let  $A = USV^{\top}$ , where U is  $m \times n$  orthonormal, S is  $n \times n$  diagonal with descending order, and  $V^{\top}$  is  $n \times n$  also orthonormal.
- From orthonormality of U, V follows that  $\|\mathbf{USV}^{\top}\mathbf{h}\| = \|\mathbf{SV}^{\top}\mathbf{h}\|$  and  $\|\mathbf{V}^{\top}\mathbf{h}\| = \|\mathbf{h}\|$ .
- Substitute  $\mathbf{y} = \mathbf{V}^{\top}\mathbf{h}$ . Now, we minimize  $\|\mathbf{S}\mathbf{y}\|$  subject to  $\|\mathbf{y}\| = 1$ .



- Let  $A = USV^{\top}$ , where U is  $m \times n$  orthonormal, S is  $n \times n$  diagonal with descending order, and  $V^{\top}$  is  $n \times n$  also orthonormal.
- From orthonormality of U, V follows that  $\|\mathbf{USV}^{\top}\mathbf{h}\| = \|\mathbf{SV}^{\top}\mathbf{h}\|$  and  $\|\mathbf{V}^{\top}\mathbf{h}\| = \|\mathbf{h}\|$ .
- Substitute  $\mathbf{y} = \mathbf{V}^{\top}\mathbf{h}$ . Now, we minimize  $\|\mathbf{S}\mathbf{y}\|$  subject to  $\|\mathbf{y}\| = 1$ .
- Remember that S is diagonal and the elements are sorted descendently. Than, it is clear that  $\mathbf{y} = [0, 0, \dots, 1]^{\top}$ .





- From orthonormality of U, V follows that  $\|\mathbf{U}\mathbf{S}\mathbf{V}^{\top}\mathbf{h}\| = \|\mathbf{S}\mathbf{V}^{\top}\mathbf{h}\|$  and  $\|\mathbf{V}^{\top}\mathbf{h}\| = \|\mathbf{h}\|$ .
- Substitute  $\mathbf{y} = \mathbf{V}^{\top}\mathbf{h}$ . Now, we minimize  $\|\mathbf{S}\mathbf{y}\|$  subject to  $\|\mathbf{y}\| = 1$ .
- Remember that S is diagonal and the elements are sorted descendently. Than, it is clear that  $\mathbf{y} = [0, 0, \dots, 1]^{\top}$ .
- From substitution we know that h = Vy from which follows that sought h is the last column of the matrix V.

- Richard Hartley and Andrew Zisserman, Multiple View Geometry in computer vision, Cambridge University Press, 2003 (2nd edition), [Appendix A5]
- Gene H. Golub and Charles F. Van Loan, Matrix Computation, John Hopkins University Press, 1996 (3rd edition).
- Eric W. Weisstein. Lagrange Multiplier. From MathWorld–A Wolfram Web Resource.
  - http://mathworld.wolfram.com/LagrangeMultiplier.html
- Eric W. Weisstein. Singular Value Decomposition. From MathWorld–A Wolfram Web Resource.
  - http://mathworld.wolfram.com/SingularValueDecomposition.html