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 $\mathbf{h}$ satisfying

$$\mathtt{A}\mathbf{h} = \mathbf{0} \,,$$

where $\mathtt{A}$ is $m \times n$ matrix, and $\mathbf{0}$ is $n \times 1$ zero vector. Assume $m \geq n$, and $\mathrm{rank}(\mathtt{A}) = n$. 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 $\mathbf{h}$ that minimizes $\|\mathtt{A}\mathbf{h}\|$ subject to $\|\mathbf{h}\| = 1$.

Derivation I — Lagrange multipliers

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

Derivation I — Lagrange multipliers

- $\mathbf{h} = \mathrm{argmin}_h \|\mathtt{A}\mathbf{h}\|$ 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 $\frac{\partial}{\partial \mathbf{h}} \left(\mathbf{h}^\top \mathtt{A}^\top \mathtt{A}\mathbf{h} + \lambda(1 - \mathbf{h}^\top \mathbf{h}) \right) = 0\,.$

Derivation I — Lagrange multipliers

- $\mathbf{h} = \operatorname{argmin}_h \|\mathtt{A}\mathbf{h}\|$ 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 $\frac{\partial}{\partial \mathbf{h}} \left(\mathbf{h}^\top \mathtt{A}^\top \mathtt{A}\mathbf{h} + \lambda(1 - \mathbf{h}^\top \mathbf{h}) \right) = 0\,.$
- We derive: $2\mathtt{A}^\top \mathtt{A}\mathbf{h} - 2\lambda\mathbf{h} = 0$.

Derivation I — Lagrange multipliers

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

Derivation I — Lagrange multipliers

- $\mathbf{h} = \mathrm{argmin}_h \|\mathtt{A}\mathbf{h}\|$ 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 $\frac{\partial}{\partial \mathbf{h}} \left(\mathbf{h}^\top \mathtt{A}^\top \mathtt{A}\mathbf{h} + \lambda(1 - \mathbf{h}^\top \mathbf{h}) \right) = 0\,.$
- We derive: $2\mathtt{A}^\top \mathtt{A}\mathbf{h} - 2\lambda\mathbf{h} = 0$.
- After some manipulation we end up with: $(\mathtt{A}^\top \mathtt{A} - \lambda\mathtt{E})\mathbf{h} = 0$ which is the characteristic equation. Hence, we know that $\mathbf{h}$ is an eigenvector of $(\mathtt{A}^\top \mathtt{A})$ and λ is an eigenvalue.
- The least-squares error is $e = \mathbf{h}^\top \mathtt{A}^\top \mathtt{A}\mathbf{h} = \mathbf{h}^\top \lambda\mathbf{h}$.

Derivation I — Lagrange multipliers

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

Derivation II — SVD

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

Derivation II — SVD

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

Derivation II — SVD

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

Derivation II — SVD

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

Derivation II — SVD

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

Further reading

- 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