Parametric Fitting

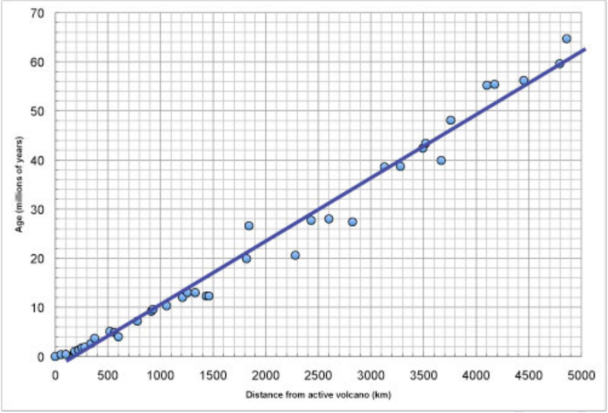
Kenichi Kanatani (interpreted by I. Esteban)

August 6, 2009

What's all about?

... fit geometric objects to multiple instances of another geometric object in an optimal manner ...

Example


70
60
50
40
30
20
10
0
Age (millions of years)
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Distance from active volcano (km)

Step by Step

- Fitting as Maximum Likelihood

Step by Step

- Fitting as Maximum Likelihood
- Obtain covariance of estimation and residual

Step by Step

- Fitting as Maximum Likelihood
- Obtain covariance of estimation and residual
- Hypothesis testing

Step by Step

- Fitting as Maximum Likelihood
- Obtain covariance of estimation and residual
- Hypothesis testing
- Examples

Object, data and relationship

- We want to fit an object given some data
- Let u be the vector that represents the object (n'-dim manifold - parameter space U)
- Let $a_1 ... a_N$ be N vectors of the same object (m'-dim manifold - data space A)
- All a_α satisfy the same relation with u
- We want to find u optimally wrt the relation

Noise

- a_α is observed in the presence of noise
  $a_\alpha = \bar{a}_\alpha + \Delta a_\alpha, \alpha = 1, ..., N.$
- Δa_α is independent, zero mean and with covariance $\bar{V}[a_\alpha]$
- To a first approximation, Δa_α lives in the tangent space $T_{\bar{a}_\alpha}(A)$

Constraints

- The true value $\bar{a}_\alpha$ and u are related by L constraints $F^{(k)}(\bar{a}_\alpha, u) = 0, k = 1, ..., L.$
- This is called the hypothesis and its assumed to be nonsingular (p.132)
- The *rank* of the hypothesis is the number of independent equations
- The *rank* is the codimension of S (manifold defined by the L eqs.)
- S is called the *geometric model* of the hypothesis

Singularity

- Linear subspace of the TRUE value
  $\bar{V}_\alpha = \{P^A_{\bar{a}_\alpha} \nabla_a F^{(1)}(\bar{a}_\alpha, u), ..., P^A_{\bar{a}_\alpha} \nabla_a F^{(L)}(\bar{a}_\alpha, u)\}_L \in R^m$
- Linear subspace of the MEASURED value
  $V_\alpha = \{P^A_{a_\alpha} \nabla_a F^{(1)}(a_\alpha, u), ..., P^A_{a_\alpha} \nabla_a F^{(L)}(a_\alpha, u)\}_L \in R^m$
- In general, the rank of the hypothesis r coincides with the dimension l of the linear subspace $\bar{V}_\alpha$
- If $l < r$ then a_α is a *singular datum*
- If the dimension of V_α is larger than the dimesion of $\bar{V}_\alpha$ the hypothesis is degenerate

Example??

- Image points
- Directions in space
- Motion estimation??

Correction

- Assume a value for u (what value?)
- Correct (optimally) data to satisfy the hypothesis
  - Find Δa_α such that $\bar{a}_\alpha = a_\alpha - \Delta a_\alpha$
  - This is the optimization:
    $J_\alpha = (\Delta a_\alpha, \bar{V}[a_\alpha]^{-} \Delta a_\alpha) \rightarrow min$
- Solution is given by:

$$\Delta a_\alpha = \bar{V}[a_\alpha] \sum_{k,l=1}^{L} \bar{W}_\alpha^{(kl)}(u) F^{(k)}(a_\alpha, u) \nabla a F^{(l)}(\bar{a}_\alpha, u)$$

- The residual $\hat{J}_\alpha$ is given by substituting one into another

Estimation

- The probability density for all Δa_α is:
$$\left(\prod_{\beta=1}^{N} \frac{1}{\sqrt{(2\pi)^{r_\beta} |\bar{V}[a_\beta]|_+}}\right) e^{-\sum_{\alpha=1}^{N}(\Delta a_\alpha, \bar{V}[a_\alpha]^- \Delta a_\alpha)/2}$$
- This is the likelihood of the observed values Δa_α ??
- Given the residual $\hat{J}_\alpha$ this takes the form:
$$\left(\prod_{\beta=1}^{N} \frac{1}{\sqrt{(2\pi)^{r_\beta} |\bar{V}[a_\beta]|_+}}\right) e^{-\sum_{\alpha=1}^{N} \hat{J}_\alpha/2}$$

Estimation 2

- We now want to find u that maximizes that likeihood (or minimizes the sum of residuals)

$$\bar{J}[u] = \sum_{\alpha=1}^{N} \hat{J}_\alpha \rightarrow min$$

- This takes the full form:

$$\bar{J}[u] = \sum_{\alpha=1}^{N} \sum_{k,l=1}^{L} \bar{W}_\alpha^{(kl)}(u) F^{(k)}(a_\alpha, u) F^{(l)}(a_\alpha, u) \rightarrow min$$

Practical Considerations

- $\bar{W}_\alpha^{(kl)}$ (pseudo inverse of V_α) depends on the true value $\bar{a}_\alpha$
- We approximate:
  $\bar{W}_\alpha^{(kl)} \approx \left((\nabla a F^{(k)}(a_\alpha, u), V[a_\alpha]\nabla F^{(l)}(a_\alpha, u)\right)_r^-$
- Thus:
  $$J[u] = \sum_{\alpha=1}^{N} \sum_{k,l=1}^{L} W_\alpha^{(kl)}(u) F^{(k)}(a_\alpha, u) F^{(l)}(a_\alpha, u) \rightarrow min$$
- Yields the optimal estimate $\hat{u}$ by numerical computation

Re-cap

- We define the optimal estimation as maximum likelihood
- We obtain an optimal estimate by approximating the true covariance with the measured covariance
- The optimal estimate $\hat{u}$ is a random variable since it was obtained from noisy data
- We now study its behavior

Some facts and definitions

- We define $\bar{u}$ as the true value and u the random variable
- $\bar{u}$ satisfies $F^{(k)}(\bar{a}_\alpha, u) = 0, k = 1, ..., L$
- Since its a random variable, its disturbed by noise: $u = \bar{u} + \Delta u$
- Δu is to a first approx contained in the tangent space

Covariance of $\hat{u}$

- We start by introducing the random variables a_α and u in the constraints:
  $F^{(k)}(a_\alpha, u) = (\nabla a\bar{F}_\alpha^{(k)}, \Delta a_\alpha) + (\nabla u\bar{F}_\alpha^{(k)}, \Delta u_\alpha) + O(\Delta a_\alpha, \Delta u)^2$
- Also:
  $\bar{W}_\alpha^{(kl)}(u) = \bar{W}_\alpha^{(kl)}(\bar{u}) + O(\Delta u)$
- After some magical math, we finally obtain:
  $$\bar{V}[\hat{u}] = \left(\sum_{\alpha=1}^{N} \sum_{k,l=1}^{L} \bar{W}_\alpha^{(kl)}(\bar{u})(P_{\bar{u}}^{U}\nabla u\bar{F}_\alpha^{(k)})(P_{\bar{u}}^{U}\nabla u\bar{F}_\alpha^{(l)})^T \right)^-$$

Approximation

- Since the true value $\bar{u}$ is used, we cannot compute it
- We approximate using the optimal estimate $\hat{u}$ and the corrected data $\hat{a}_\alpha = a_\alpha - \Delta a_\alpha$
- Alternatively, one approximation can be made using the measured data and not the corrected one.

Re-cap

- The estimation is based on the *hypothesis* that the data $\{a_\alpha\}$ are random deviations from the true data $\{\bar{a}_\alpha\}$
- The true data satisfies the constraints $F^{(k)}$
- Minimizing the sum of residuals means choosing $\hat{u}$ so that the hypothesis is most likely

Hypothesis testing

- If the hypothesis is correct, the residual $\bar{J}[\hat{u}]$ should be zero for the true values
- This is generally NOT the case
- The bigger the residual, the lest likely the hypothesis is correct
- If the residual is much larger than expected according to the noise in the data $\{a_\alpha\}$ the hypothesis can be rejected
- To formalize this, we assume Gaussian noise in the data

Strong hypothesis

- We want to reject a strong hypothesis
- A strong hypothesis is based on the residual of the TRUE value $\bar{u}$
- So, we consider the residual $\bar{J}[\bar{u}]$ and we let $\Delta u = 0$
- This is given by:
$$\bar{J}[\bar{u}] = \sum_{\alpha=1}^{N} \sum_{k,l=1}^{L} \bar{W}_{\alpha}^{(kl)}(\bar{u})(\nabla a \bar{F}_{\alpha}^{(k)}, \Delta a_{\alpha})(\nabla a \bar{F}_{\alpha}^{(l)}, \Delta a_{\alpha})$$
- We now re-write it using a random variable of mean 0 (the e_{α} vector) and it covaraince as:
$$\bar{J}[\bar{u}] = \sum_{\alpha=1}^{N} (e_{\alpha}, V[e_{\alpha}]^{-} e_{\alpha})$$

Strong hypothesis 2

- The rank of the covariance of e_α is the same as the hypothesis and each e_α is a random variable (Gaussian and independent)
- So $\bar{J}[\bar{u}]$ is a χ^2 variable, so we apply the ol' rejection method
- The hypothesis can be rejected with significance level $a\%$ if: $\bar{J}[\bar{u}] > \chi^2_{rN,a}$
- Since $\bar{J}[u]$ requires the true data, it is approximated with the residual given the measured data $J[u]$

Weak hypothesis

- The same idea as before, but now we use the residual for the optimal estimate $\bar{J}[\hat{u}]$
- To a first approximation is:
  $\bar{J}[\hat{u}] = \bar{J}[\bar{u}] - (\Delta u, \bar{V}[\hat{u}]^{-}\Delta u)$
- The first part is a χ^2 variable.... and also the second part
- The expectation and variance are LOWER than the ones for the strong hypothesis:
  $E[\bar{J}[\bar{u}]] = rN, V[\bar{J}[\bar{u}]] = 2rN$
  $E[\bar{J}[\hat{u}]] = rN - n', V[\bar{J}[\hat{u}]] = 2(rN - n')$

Weak hypothesis 2

- The residual for the optimal estimate is (whp) smaller than for $\bar{u}$
- This is because its obtained minimizing the residual
- This analysis can be used to test that the constraints $F^{(k)}$ are satiesfied by some value u
- The hypothesis is rejected with significance value a% if $\bar{J}[\hat{u}] > \chi^2_{rN-n',a}$
- The same approximation is made for the residual as before

Noise level

- If a covariance matrix y multiplied by a constant c...
- ... the pseudo inverse is multiplied by $1/c$
- This does not affect the value that minimizes the residual (only the residual scale)
- The covariance is expressed as:
  $V[a_\alpha] = \epsilon^2 V_0[a_\alpha]$

Noise level 2

- In practical problems, V_0 can be predicted, but not ϵ
- We can first estimate $\hat{u}$...
- and later estimate the noise level:
  $\hat{\epsilon}^2 = \frac{J_0[\hat{u}]}{rN-n'}$
- The weak hypothesis can be re-writen as:
  $\frac{\hat{\epsilon}^2}{\epsilon^2} > \frac{\chi^2_{rN-n',a}}{rN-n'}$

Matlab example

The covariance

- K. assumes that you posses the covariance of the original data... but can you compute it?
- Example:
  - Given image correspondences, compute camera motion
  - Then triangulate and obtain 3D points
  - Fit some geometric object (line, plane, whatever)
- How does the error propagate thorough this?