Umeyama-Kabsch Algorithm [1]

Aligning point patterns

Davide Malvezzi

2022 October

1.1 Umeyama-Kabsch algorithm

Given two sets of n-dimensional points x_i and y_i we want to find the scaling c, rotation R and translation t to align the two sets and which minimize

$$e^2(c, R, t) = \frac{1}{n}\sum_{i=1}^{n}\|y_i - (cRx_i + t)\|^2 \tag{1.1}$$

1.1.1 Least-square estimation for rotation only

Let's start considering only the estimation of the rotation matrix.
Given A and B, mxn matrix, whose columns are points from the sets y_i and x_i respectively and R mxm rotation matrix, we seek to minimize

$$\min_R \|A - RB\|^2 \tag{1.2}$$

under the constraints

$$R^T R = I \tag{1.3}$$
$$det(R) = 1 \tag{1.4}$$

which define R as a valid rotation matrix.
The corresponding Lagrangian is

$$F = \|A - RB\|^2 + \mathrm{tr}(L(R^T R - I)) + g(det(R) - 1) \tag{1.5}$$

Note 1.1: What is tr$(L(R^T R - I))$?

Let R = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$
then $R^T R = 1$ becomes

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0 \tag{1.6}$$
$$\begin{bmatrix} f-1 & g \\ g & h-1 \end{bmatrix} = 0 \tag{1.7}$$
$$\tag{1.8}$$

Therefore, we have 3 constraints, $f = 1$, $h = 1$ and $g = 0$ and the corrisponding Lagrangian multipliers α, β and γ.
Then L is a symmetric matrix of Lagrangian multipliers

$$L = \begin{bmatrix} \alpha & \gamma \\ \gamma & \beta \end{bmatrix} \tag{1.9}$$

and $L(R^T R - I)$ is

$$\begin{bmatrix} \alpha & \gamma \\ \gamma & \beta \end{bmatrix}\begin{bmatrix} f-1 & g \\ g & h-1 \end{bmatrix} = \begin{bmatrix} \alpha(f-1) + \gamma g & \alpha g + \gamma(h-1) \\ \gamma(f-1) + \beta g & \gamma g + \beta(h-1) \end{bmatrix} \tag{1.10}$$

But we just need the main diagonal for the Lagrangian, so we apply the tr operator.

1.1.2 Derivatives

Deriving F we obtain

$$\frac{\partial F}{\partial R} = -2AB^T + 2RBB^T + 2RL + gR = 0 \tag{1.11}$$

$$\frac{\partial F}{\partial L} = R^T R - I = 0 \tag{1.12}$$

$$\frac{\partial F}{\partial g} = det(R) - 1 = 0 \tag{1.13}$$

Note 1.2: Matrix norm

Note that

$$\|A\|^2 = \mathrm{tr}(A^T A)$$

Note 1.3: Matrix derivative

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}] = \underline{I} \tag{A.1}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\underline{X}] = \underline{A}' \tag{A.2}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\underline{X}'] = \underline{A} \tag{A.3}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\underline{X}\underline{B}] = \underline{A}'\underline{B}' \tag{A.4}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\underline{X}'\underline{B}] = \underline{B}\underline{A} \tag{A.5}$$

$$\frac{\partial}{\partial \underline{X}'} \mathrm{tr}[\underline{A}\underline{X}] = \underline{A} \tag{A.6}$$

$$\frac{\partial}{\partial \underline{X}'} \mathrm{tr}[\underline{A}\underline{X}'] = \underline{A}' \tag{A.7}$$

$$\frac{\partial}{\partial \underline{X}'} \mathrm{tr}[\underline{A}\underline{X}\underline{B}] = \underline{B}\underline{A} \tag{A.8}$$

$$\frac{\partial}{\partial \underline{X}'} \mathrm{tr}[\underline{A}\underline{X}'\underline{B}] = \underline{A}'\underline{B}' \tag{A.9}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}\underline{X}] = 2\underline{X}' \tag{A.10}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}\underline{X}'] = 2\underline{X} \tag{A.11}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}^n] = n(\underline{X}^{n-1})' \tag{A.12}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\underline{X}^n] = \Big(\sum_{i=0}^{n-1} \underline{X}^i \underline{A}\underline{X}^{n-1-i}\Big)' \tag{A.13}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\underline{X}\underline{B}\underline{X}] = \underline{A}'\underline{X}'\underline{B}' + \underline{B}'\underline{X}'\underline{A}' \tag{A.14}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\underline{X}\underline{B}\underline{X}'] = \underline{A}'\underline{X}\underline{B}' + \underline{A}\underline{X}\underline{B} \tag{A.15}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[e^{\underline{X}}] = e^{\underline{X}'} \tag{A.16}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}^{-1}] = -(\underline{X}^{-1}\underline{X}^{-1})' = -(\underline{X}^{-2})' \tag{A.17}$$

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\underline{X}^{-1}\underline{B}] = -(\underline{X}^{-1}\underline{B}\underline{A}\underline{X}^{-1})' \tag{A.18}$$

$$\frac{\partial}{\partial \underline{X}} \det[\underline{X}] = (\det[\underline{X}])(\underline{X}^{-1})' \tag{A.19}$$

$$\frac{\partial}{\partial \underline{X}} \log\det[\underline{X}] = (\underline{X}^{-1})' \tag{A.20}$$

$$\frac{\partial}{\partial \underline{X}} \det[\underline{A}\underline{X}\underline{B}] = (\det[\underline{A}\underline{X}\underline{B}])(\underline{X}^{-1})' \tag{A.21}$$

$$\frac{\partial}{\partial \underline{X}} \det[\underline{X}'] = \frac{\partial}{\partial \underline{X}} \det[\underline{X}] = (\det[\underline{X}])(\underline{X}^{-1})' \tag{A.22}$$

$$\frac{\partial}{\partial \underline{X}} \det[\underline{X}^n] = n(\det[\underline{X}])^n(\underline{X}^{-1})' \tag{A.23}$$

1.1.3 Decomposing L'

From the first derivative

$$RL' = AB^T \quad \text{where} \quad L' = BB^T + L + \frac{1}{2}gI \tag{1.14}$$

Note L' is symmetric because sum of symmetric matrices.
Multiplying each side by its transpose

$$(RL')^T RL' = (AB^T)^T AB^T \tag{1.15}$$
$$L'^T R^T RL' = BA^T AB^T \tag{1.16}$$
$$L'^T L = BA^T AB^T \tag{1.17}$$
$$L'^2 = BA^T AB^T \tag{1.18}$$

Let $AB^T = UDV^T$ be the SVD decomposition, then

$$L'^2 = BA^T AB^T \tag{1.19}$$
$$L'^2 = VDU^T UDV^T \tag{1.20}$$
$$L'^2 = VD^2V^T \tag{1.21}$$

Given $S = diag(s_1, ..., s_n)$, with $s_i = 1$ or $s_i = -1$, the matrix L' can be written then as

$$L' = VDSV^T \tag{1.22}$$

Proof:

$$L'^2 = L'L' = VDSV^T VDSV^T = VD^2V^T \tag{1.23}$$

😄

1.1.4 Determinant equivalence

Now we have

$$det(L') = det(VDSV^T) = det(V)det(D)det(S)det(V^T) = det(D)det(S) \tag{1.24}$$
$$det(L') = det(R^T AB^T) = det(R^T)det(AB^T) = det(AB^T) \tag{1.25}$$

Therefore

$$det(D)det(S) = det(AB^T) \tag{1.26}$$

Since singular value are always positive, $det(D) > 0$ then $det(S)$ must be equal to 1 or -1 depending on the sign of $det(AB^T)$

1.1.5 Replacing into F

We can rewrite F with the results obtained

$$F = \|A - RB\|^2 = \|A\|^2 + \|B\|^2 - 2\text{tr}(AB^T R^T) \tag{1.27}$$
$$= \|A\|^2 + \|B\|^2 - 2\text{tr}(R^T AB^T) \quad \text{\#trace of matrix product is indipendent on order} \tag{1.28}$$
$$= \|A\|^2 + \|B\|^2 - 2\text{tr}(L') \tag{1.29}$$
$$= \|A\|^2 + \|B\|^2 - 2\text{tr}(VDSV^T) \tag{1.30}$$
$$= \|A\|^2 + \|B\|^2 - 2\text{tr}(V^T VDS) \quad \text{\#trace of matrix product is indipendent on order} \tag{1.31}$$
$$= \|A\|^2 + \|B\|^2 - 2\text{tr}(DS) \tag{1.32}$$
$$= \|A\|^2 + \|B\|^2 - 2(d_1s_1 + ... + d_ms_m) \tag{1.33}$$

Being the singular value in D always positive, the minimum of F is obtained when all $s_i = 1$ if $det(AB^T) \geqslant 0$ otherwise if $det(AB^T) < 0$ we must set $s_m = -1$ to satisfy the determinant equivalence.

1.1.6 Getting optimal R

When AB^T is full rank therefore L'^{-1} exists, so

$$R = AB^T L'^{-1} = UDV^T VD^{-1}SV^T = USV^T \tag{1.34}$$

When $rank(AB^T) = m - 1$ we have

$$RL' = AB^T \tag{1.35}$$
$$RVDSV^T = UDV^T \tag{1.36}$$
$$RVDS = UD \tag{1.37}$$
$$U^T RVDS = D \tag{1.38}$$
$$QD = D \tag{1.39}$$

where $DS = D$ because $s_i = 1 \quad i = 1...m-1$ and $d_m = 0$. The new matrix $Q = U^T RV$ is orthogonal by construction, so the last equation is true only if each column $q_i = e_i \quad i = 1...m-1$. For the last column q_m possibile solutions are e_m or $-e_m$, this depends on the determinat value

$$det(Q) = det(U^T)det(R)det(V) = det(U)det(V) \tag{1.40}$$

so if $det(U)det(V) = 1$ then $det(Q) = 1$ so $q_m = e_m$ otherwise $q_m = -e_m$ In conclusion

$$R = UQV^T \tag{1.41}$$

1.1.7 Least-square estimation for c, R and t

Let's now consider the original problem

$$e^2(c, R, t) = \frac{1}{n}\sum_{i=1}^{n} \|y_i - (cRx_i + t)\|^2 = \frac{1}{n}\|Y - cRX - th^T\|^2 \tag{1.42}$$

Definition 1.1.1: Normalization matrix

We define a nxn normalization matrix K as

$$K = I - \frac{1}{n}hh^T \quad \text{where} \quad h = (1, \dots, 1)^T \tag{1.43}$$

with the following properties

$$K = K^T = K^2 \tag{1.44}$$

Definition 1.1.2

Then we can define

$$\mu_x = \frac{1}{n} Xh \tag{1.45}$$

$$\mu_y = \frac{1}{n} Yh \tag{1.46}$$

$$\sigma_x^2 = \frac{1}{n} \|XK\|^2 \tag{1.47}$$

$$\sigma_y^2 = \frac{1}{n} \|YK\|^2 \tag{1.48}$$

$$\Sigma_{xy} = \frac{1}{n} YKX^T \tag{1.49}$$

$$X = XK + \frac{1}{n} Xhh^T \tag{1.50}$$

$$Y = YK + \frac{1}{n} Yhh^T \tag{1.51}$$

and substitute in the loss function

$$\begin{aligned}
e^2(c, R, t) &= \frac{1}{n} \|Y - cRX - th^T\|^2 && (1.52)\\
&= \frac{1}{n} \|YK + \frac{1}{n} Yhh^T - cRXK - \frac{c}{n} RXhh^T - th^T\|^2 && (1.53)\\
&= \frac{1}{n} \|YK - cRXK + \left(\frac{1}{n} Yh - \frac{c}{n} RXh - t\right) h^T\|^2 && (1.54)\\
&= \frac{1}{n} \|YK - cRXK + t'h^T\|^2 && (1.55)\\
&= \frac{1}{n} \left(\|YK - cRXK\|^2 + \|t'h^T\|^2 - 2\text{tr}(K(Y^T - cX^T R^T) t'h^T)\right) && (1.56)
\end{aligned}$$

Let's analyze each single term of the loss.

First term

If we define $A = YK$ and $B = XK$, the optimal R of the first term is given by the solution of 1.2.
Also note that

$$AB^T = YKK^T X^T = YKX^T = n\Sigma_{xy} \tag{1.57}$$

Note 1.4: SVD decomposition of a constant times a matrix

Given a matrix A and the matrix cA, their SVD decomposition are equivalent except the singular value of the second are the singular of the first matrix times c

So, given the SVD decomposition of $\Sigma_{xy} = UDV^T$, the optimal $R = USV^T$.
Then the first term can be decomposed as

$$\begin{aligned}
\frac{1}{n} \|YK - cRXK\|^2 &= \frac{1}{n} \left(\|YK\|^2 + \|cRXK\|^2 - 2\text{tr}((YK)^T (cRXK))\right) && (1.58)\\
&= \frac{1}{n} \left(\|YK\|^2 + \|cXK\|^2 - 2\text{tr}((YK)(cRXK)^T)\right) && (1.59)\\
&= \frac{1}{n} \left(\|YK\|^2 + \|cXK\|^2 - 2\text{tr}(cYKK^T X^T R^T)\right) && (1.60)\\
&= \frac{1}{n} \left(\|YK\|^2 + \|cXK\|^2 - 2\text{tr}(cYKX^T R^T)\right) && (1.61)
\end{aligned}$$

Let UDV^T be the SVD decomposition of $\Sigma_{xy} = \frac{1}{n} YKX^T$, then

$$YK(cXK)^T = cYKK^T X^T = cYKKX^T = cYK^2 X^T = cYKX^T = cnUDV^T \tag{1.62}$$

Substituting back $R = USV^T$ and $cYKX^T = cnUDV^T$

$$
\begin{aligned}
\frac{1}{n}\|YK - cRXK\|^2 &= \frac{1}{n}\left(\|YK\|^2 + \|cXK\|^2 - 2\mathrm{tr}(YKX^TR^T)\right) && (1.63)\\
&= \frac{1}{n}\left(\|YK\|^2 + \|cXK\|^2 - 2\mathrm{tr}(cnUDV^T(USV^T)^T)\right) && (1.64)\\
&= \frac{1}{n}\left(\|YK\|^2 + \|cXK\|^2 - 2\mathrm{tr}(cnUDV^TVS^TU^T)\right) && (1.65)\\
&= \frac{1}{n}\left(\|YK\|^2 + \|cXK\|^2 - 2\mathrm{tr}(cnUDS^TU^T)\right) && (1.66)\\
&= \frac{1}{n}\left(\|YK\|^2 + \|cXK\|^2 - 2\mathrm{tr}(cnDS)\right) && (1.67)\\
&= \sigma_y^2 + c^2\sigma_x^2 - 2c\mathrm{tr}(DS) && (1.68)
\end{aligned}
$$

In conclusion we have a quadratic equation in c, so

$$c = \frac{\mathrm{tr}(DS)}{\sigma_x^2} \tag{1.69}$$

Second term

The second term can be rewritten as

$$\|t'h^T\|^2 = \|t'\|^2\|h^T\|^2 = n\|t'\|^2 \tag{1.70}$$

So to minimize the loss t' must be 0

$$
\begin{aligned}
t' &= 0 && (1.71)\\
\frac{1}{n}Yh - \frac{c}{n}RXh - t &= 0 && (1.72)\\
t &= \frac{1}{n}Yh - \frac{c}{n}RXh && (1.73)\\
t &= \mu_y - cR\mu_x && (1.74)
\end{aligned}
$$

Third term

Observing that $h^TK = h^T(I - \frac{1}{n}hh^T) = h^T - \frac{1}{n}h^Thh^T = h^T - h^T = 0$, so the third term is always equal to 0

$$\mathrm{tr}(K(YT - cXTRT)t'h^T) = \mathrm{tr}(K(YT - cXTRT)t'h^TKK^{-1}) = 0 \tag{1.75}$$

1.1.8 Algorithm

Algorithm 1 Umeyama-Kabsch Algorithm

Require: X mxn matrix
Require: Y mxn matrix
  μ_x = mean(X)
  μ_y = mean(Y)
  σ_x^2 = var(X)
  Σ_{xy} = cov(X, Y)
  S = eye(n)
  U, D, V^T = svd(Σ_{xy})
  if det(U) det(V^T) < 0 **then**
    S[n, n] = -1
  end if
  c = trace(DS) / σ_x^2
  $R = USV^T$
  $t = \mu_y - cR\mu_x$
  return c, R, t

Bibliography

[1] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991.