Lecture 11. Square root SAM ($\sqrt{\text{SAM}}$)

Gonzalo Ferrer

1 March 2021

1 Chi squared error

$$\chi^2 = \sum \|\cdot\|^2_{\Sigma_i} + \|\cdot\|^2_{\Sigma_k} = \|A\delta - b\|_2^2$$

If χ^2 has converged ($\delta = 0$), then

$$\chi^2\big|_{\delta=0} = b^T b$$

1.1 Information matrix ($A^T A = \Lambda$)

SAM is a MAP estimator of $x_{0:t_i}$

$$\arg\max_{\theta} P(\mathcal{X}, \mathcal{M}, \mathcal{Z}, \mathcal{U}) \xrightarrow{(-\log\,,\ \text{linearisation})} \arg\min_{\delta} \|A\delta - b\|_2^2$$

In fact, all these factors express a distribution as well.

$$\begin{aligned}\|A\delta - b\|_2^2 = (A\delta - b)^T(A\delta - b) = \delta^T A^T A\delta - \delta^T A^T b - b^T A\delta + b^T b \overset{\text{if } b=A\mu}{=} \\ \delta^T A^T A\delta - 2\delta^T A^T A\mu + \mu^T A^T A\mu = \\ (\delta - \mu)^T \underbrace{A^T A}_{\Lambda}(\delta - \mu)\end{aligned}$$

1.2 Normal equation

$$\begin{aligned} A\delta &= b \Rightarrow (\delta = A^{-1} \cdot b) \\ A^T A\delta &= A^T b \\ \delta &= (A^T A)^{-1} A^T b. \qquad \text{complexity of this: } \mathcal{O}(n^3)\end{aligned}$$

A is space $\rightarrow$ exploit by SOTA Linear algebra.

1.3 Cholesky factorisation

$\Lambda = A^T A = L \cdot L^T = R^T R$, where R is an upper-triangular matrix (can be a lower-triang.)

$$\begin{aligned} A^T A\delta &= A^T b \\ &\Downarrow \text{(cholesky)} \\ R^T R\delta &= A^T b \qquad \text{(Squared root method)}\end{aligned}$$

$$\left[\begin{aligned} R^T \cdot y &= A^T b \\ R \cdot \delta &= y \end{aligned}\right.$$

Equations above solved efficiency by back-substitution

Example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 5 & 7 & 0 \\ 6 & -7 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 5 \end{bmatrix}$$

Solving:

1. $y_1 = 2$
2. $5 \cdot 2 + 7y_2 = 5 \Rightarrow y_2 = -\dfrac{5}{7}$
3. $6y_1 - 7y_2 + 3y_3 = 5$
   $6 \cdot 2 - 7(-\frac{5}{7}) + 3y_3 \Rightarrow y_3 = \dfrac{5 - 12 - 5}{3} = 4$

Cholesky factorisation regiven to solve 2 systems by back-substitution

1.4 QR factorisation

$$Q^T A = \begin{bmatrix} R \\ 0 \end{bmatrix}$$
$$Q^T b = \begin{bmatrix} d \\ e \end{bmatrix}$$

Q is orthonormal matrix (square)
R is upper triangular matrix.

$$\|A\delta - b\|_2^2 = \|Q^T A\delta - Q^T b\|_2^2 = \| \begin{bmatrix} R \\ 0 \end{bmatrix} \delta - \begin{bmatrix} d \\ e \end{bmatrix} \|_2^2 = \\ \|R\delta - d\|_2^2 + \|e\|_2^2$$

$\boxed{\mathrm{R}\delta = d}$ solved by 1 back-substitution

Note: no need to calculate $\Lambda = A^T A$

Is R the same as for Cholesky?

$$\Lambda = A^T A = (QR)^T QR = R^T \underbrace{Q^T Q}_{I} R = R^T R$$

1.5 Schur-complement for factorisation (g2o, Kumarle'2011)

$$\Lambda = A^T A = \begin{bmatrix} \Lambda_x & \Lambda_{xm} \\ \Lambda_{mx} & \Lambda_m \end{bmatrix}, \delta = \begin{bmatrix} \delta_x \\ \delta_m \end{bmatrix}, A^T b = \begin{bmatrix} b_x \\ b_m \end{bmatrix}$$

Λ_m is diagonal.

$$\text{from} \quad A^T A\delta = A^T$$

$$\begin{cases} \Lambda_x \delta_x + \Lambda_{xm}\delta_m = b_x \\ \Lambda_{mx}\delta_x + \Lambda_m \delta_m = b_m \quad (\times \Lambda_{xm}\Lambda_m^{-1}) \end{cases}$$

$$- \quad \begin{cases} \Lambda_x \delta_x + \Lambda_{xm}\delta_m & = b_x \\ \Lambda_{xm}\Lambda_m^{-1}\Lambda_{mx}\delta_x + \Lambda_{xm}\cancel{\Lambda_m^{-1}}\cancel{\Lambda_m}\delta_m & = \Lambda_{xm}\Lambda_m^{-1} b_m \end{cases}$$

$$\underbrace{(\Lambda_x - \Lambda_{xm}\Lambda_m^{-1}\Lambda_{mx})}_{\text{S, Schur complement}} \delta_x + 0 = \underbrace{b_x - \Lambda_{xm}\Lambda_m^{-1} b_m}_{b_s}$$

1. Solve $S\delta_x = b_s$ using Cholesky for instance;
2. $\Lambda_{mx}\delta_x + \Lambda_m\delta_m = b_m$
   $\delta_m = \Lambda_m^{-1}(b_m - \Lambda_{mx}\delta_x)$.

2 Ordering of nodes

Every graph has an optimal ordering of nodes.

- fewer edges when eliminating nodes (equivalent to back-substitution in linear algebra);
- fewer fill-ins in the square root factorisation.

Example:

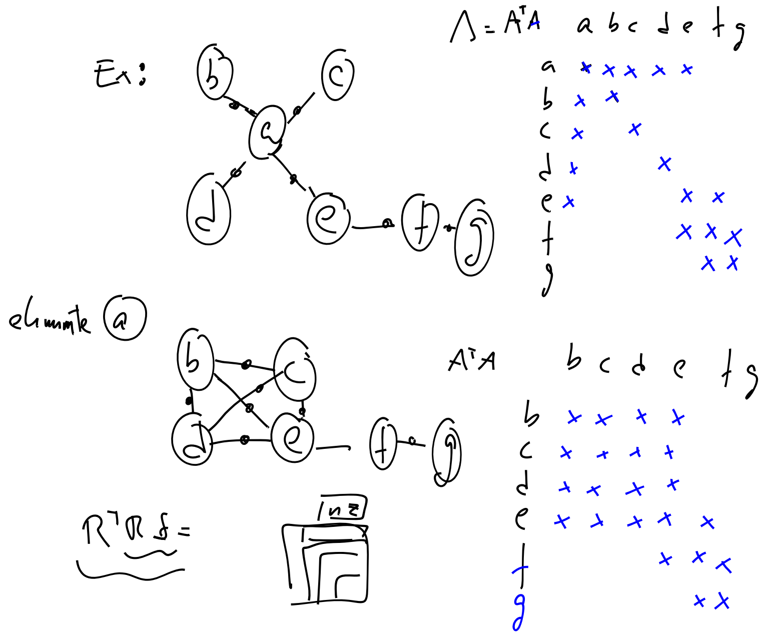

Figure 1: Ordering of nodes example

New information matrix is denser then expected. Adjacency matrix has more (factors) rows $A'_{8\times6}$ while before $A_{6\times7}$

3 Minimum order degree

Intuition: nodes with fewer edges.
Permute nodes: in order of number of edges.
Column permutation $(A^T A)$ Cholesky
Column permutation (A) QR

COLAMD heuristic for ordering nodes (Davis' 2001).
As reported in Dellaert'2006 the performed better using Cholesky and COLAMD. But it is problem dependent.