Cramér-Rao bounds for target tracking

Branko Ristic, DSTO, Australia

Int. Conf. Intelligent Sensors, Sensor Networks and Information Processing, 6 Dec. 2005
Talk Outline

- What is CRB and why do we need it?
- CRB for nonlinear filtering
- CRB for jump Markov processes
- CRB for uncertain data association
- Multi-target CRB
- Sensor allocation using CRB
- Summary
What is Cramér-Rao bound?

- CR inequality provides a lower bound on the achievable mean-square estimation error.

- The CRB for unbiased estimators mainly in use (although the formulation for biased estimators is also available);

- We distinguish two cases:
  - deterministic parameter estimation
  - stochastic parameter estimation (a.k.a. posterior CRB)

- Existence of the CR bound not guaranteed.
Some history

• The CR inequality was first stated by Ronald Fisher (1925).

• Proven by Daniel Dugué (1937).

• Harold Cramér, C. R. Rao (independently) merely re-derived the bound (1945)!

• H. Van Trees (1968) introduced the bound to a wider engineering community.
Applications of the CR bound (tracking context)

- Theoretically possible to predict the best achievable 2nd-order error performance for a target tracking problem (before you develop an algorithm);

- Aid in a tracker design: one can assess the effects of approximations embedded in tracking algorithms (by comparing RMS errors with the bound);

- Sensor management applications:
  - radar scheduling;
  - spatial deployment of sonobuoys;
  - observer trajectories (bearings-only tracking, cooperative UAVs, etc)
Definition (static case)

- Suppose $x$ is an unknown random parameter vector (dim $n_x$)

- $Z = (z_1, \ldots, z_k)$ is a vector of measurement data

- Let $\hat{x} = g(Z)$ be an unbiased estimate of $x$.

- The Cramér-Rao inequality:

  \[ C \triangleq \mathbb{E} \left\{ [g(Z) - x] [g(Z) - x]^T \right\} \geq J^{-1} \]

- $J$ is the (Fisher) information matrix with elements:

  \[ J_{ij} = -\mathbb{E} \left[ \frac{\partial^2 \ln p(x, Z)}{\partial x_i \partial x_j} \right] \quad (i, j = 1, \ldots, n_x) \]
Some properties of the bound

- Inequality $C \geq J^{-1}$ means that the difference $C - J^{-1}$ is a positive semi-definite matrix;

- Since $p(x, Z) = p(Z|x) \cdot p(x)$, the information matrix decomposed as:
  \[ J = J_z + J_p \]
  where $J_z$ represents the information obtained from the data and $J_p$ represents the prior information.

- If prior pdf $p(x)$ is a multivariate Gaussian with covariance $P_0$, then $J_p = P_0^{-1}$

- The diagonal elements of $J^{-1}$ are lower bounds of the corresponding mean-square error.
Nonlinear Filtering Problem (dynamic systems)

Notation:
• $k$ is the discrete-time index
• $x_k$ is target state vector at time $k$
• $z_{\ell k}$ is the measurement vector at time $k$ from sensor $\ell = 1, \ldots, L$
• $w_k, v_k$ are independent white processes
• $f_k(\cdot), h_k(\cdot)$ are nonlinear functions

\[
x_k = f_{k-1}(x_{k-1}) + w_{k-1}
\]

\[
z_{\ell k} = h_{k}^\ell(x_k) + v_{k}^\ell
\]

for $k = 1, 2, 3, \ldots$

The assumption is that the initial state $x_0$ has a known pdf $p(x_0)$. 
The CR bound for the Nonlinear Filtering Problem

- Research topic for about three decades:
  ⇒ an excellent review by T. H. Kerr (1989)


$$J_{k+1} = J_p(k + 1) + \sum_{\ell} J_{\ell z}(k + 1)$$

- $J_p(k + 1)$ is prior (or predicted) information matrix
- $J_{\ell z}(k + 1)$ is information matrix due to measurement from sensor $\ell = 1, \ldots, L$ at time $k$. Further on we assume $\ell = 1$ for simplicity.
The CR bound for the Nonlinear Filtering Problem (Cont’d)

- If process noise $w_k \sim \mathcal{N}(0, \Sigma_k)$, and $\Sigma_k$ non-singular, then

$$J_p(k + 1) = \Sigma_k^{-1} - \Sigma_k^{-1} \mathbb{E}\{F_k\} \left( J_k + \mathbb{E}\{F_k^T \Sigma_k^{-1} F_k\} \right)^{-1} \mathbb{E}\{F_k^T\} \Sigma_k^{-1}$$

where $F_k = \left[ \nabla_{x_k} [f_k(x_k)]^T \right]^T$ is the Jacobian of $f_k(\cdot)$.

- If measurement noise $v_k \sim \mathcal{N}(0, R_k)$, and $R_k$ non-singular, then

$$J_z(k) = \mathbb{E} \left\{ H_k^T R_k^{-1} H_k \right\}$$

where $H_k = \left[ \nabla_{x_k} [h_k(x_k)]^T \right]^T$ is the Jacobian of $h_k(\cdot)$.
Nonlinear Filtering: Deterministic case

- In the absence of process noise, i.e. \( w_k = 0 \), target state \( x_k \) is an unknown deterministic parameter (knowing \( x_0 \) we can compute \( x_k \) for any \( k \));

- The expectation operator \( E \) disappears; a simple recursive formula [Taylor, 1979]:

\[
J_{k+1} = \left( F_k^{-1} \right)^T J_k F_k^{-1} + H_{k+1}^T R_{k+1}^{-1} H_{k+1}
\]

- Observation: This is identical to the covariance matrix propagation formula for the Extended Kalman filter! There is only one difference: here we use true values of \( x_k \) to evaluate Jacobians \( F_k \) and \( H_k \).
Examples: Bearings-only tracking

- Bearings measurements collected asynchronously by distributed sensors

- Target moving with a (nearly) constant velocity (linear dynamics);
  \[
  \mathbf{x}_k = \begin{bmatrix} x_k & \dot{x}_k & y_k & \dot{y}_k \end{bmatrix}^T
  \]

- Sensors are mobile; sensor state vector is known:
  \[
  \mathbf{x}_k^\ell = \begin{bmatrix} x_k^\ell & \dot{x}_k^\ell & y_k^\ell & \dot{y}_k^\ell \end{bmatrix}^T, \quad \ell \in \{1, 2, \ldots, L\} \]
Examples: Bearings-only tracking (Cont’d)

- Measurement equation (nonlinear)

\[ z_{\ell k} = h_{\ell k}(x_k) + v_{\ell k}, \quad h_{\ell k}(x_k) = \arctan \frac{y_k - y_{\ell k}}{x_k - x_{\ell k}} \]

- \( z_{\ell k} \) is a measurement from sensor \( \ell_k \) at time \( t_k \);

- \( v_{\ell k} \) is measurement noise in sensor \( \ell_k \): zero-mean white Gaussian, with variance \( R_{\ell k} = \sigma_{\ell k}^2 \).

- Estimation problem:
  Given sensor messages \( \mathcal{M}_k = \{(t_i, x_i^{\ell_i}, z_i^{\ell_i})\} (i = 1, \ldots, k) \), estimate \( x_k \).
Examples: Bearings-only tracking (Cont’d)

Single mobile sensor (must manoeuvre to observe the target state)
Examples: Bearings-only tracking (Cont’d)

Two Mobile sensors: Sensor 1 as before; Sensor 2 reports only at: 31.6s, 47.6s, 63.6s, 79.6s, 95.6s, 111.6s
Examples: Tracking a Ballistic Object on Re-entry

- Problem:
  
  Sequential estimation of kinematic parameters (position, velocity) of a ballistic object re-entering the atmosphere

- Practical applications: Surveillance for missile defence (e.g. scud missiles)

- Problem *difficult* due to the nonlinear object dynamics;

Ballistic Object on Re-entry: Dynamics

- 1D (vertical) motion
- Only two forces act upon the object: drag (air resistance) and gravity
- Differential equations:

\[
\dot{h} = -v \\
\dot{v} = \frac{\rho(h) \cdot g \cdot v^2}{2\beta} - g
\]

where
- \( h \) - object height;
- \( v \) - object velocity;
- \( \beta \) - ballistic coefficient (depends on mass, shape, cross-sec.);
- \( \rho(h) = \gamma \cdot e^{-\eta h} \) (air density);
- \( g = 9.81 \text{m/s}^2 \)
Ballistic Object on Re-entry: Dynamics & measurements

- State vector $x_k = [h_k, v_k, \beta_k]^T$;

- Using Euler approx. with a small integration step $\tau$
  
  $$x_{k+1} = f_k(x_k) + w_k$$

  where $f_k(x_k)$ is nonlinear due to drag $D(x_k) = \frac{g \cdot \rho(x_k[1]) \cdot x_k[2]^2}{2x_k[3]}$

- Process noise: $w_k \sim \mathcal{N}(0, \Sigma)$

- Radar is measuring target height (range) every $T \geq \tau$ seconds;

- Measurement equation is linear:
  
  $$z_k = Hx_k + v_k$$

  where $H = [1 \ 0 \ 0]$ and $v_k \sim \mathcal{N}(0, R = \sigma_r^2)$.

Ref: B. Ristic, S. Arulampalam, N. Gordon, Beyond the Kalman filter, 2004 (chapter 5).
Ballistic Object on Re-entry: Trajectory

- \( h_0 = 60960 \text{ m}; \)
- \( v_0 = 3048 \text{ m/s}; \)
- \( \beta_0 = 23948 \text{ kg/ms}^2 \) (corresp. mass of 500 kg)
Ballistic Object on Re-entry: CR bound

- \( R = (200 \text{m})^2; \)

- \( \sigma_\beta = 7184 \text{ kg/ms}^2 \)
CRB for switching dynamic models

- Object motion sometimes must be modelled using more than a single dynamic model;

- Typical motion models: constant velocity, constant acceleration, coordinated turn, etc.
Switching Dynamic model

- Multiple switching linear dynamic models with additive Gaussian noise:
  \[ x_{k+1} = F_k(r_{k+1})x_k + w_k(r_{k+1}) \]

- \( r_{k+1} \) specifies the target motion model (or regime) which is in effect during the time interval \( (t_k, t_{k+1}] \);

- \( w_k(r_{k+1}) \sim \mathcal{N}(0, \Sigma_k(r_{k+1})) \);

- The evolution of motion model sequence is modelled by a time-homogeneous Markov chain with known:
  - transitional probabilities
    \[ \pi_{ij} \triangleq \mathbb{P}\{r_{k+1} = j | r_k = i\}, \quad i, j \in S \triangleq \{1, 2, \ldots, s\} \]
  - initial motion model probabilities:
    \[ p_1(i) \triangleq \mathbb{P}\{r_1 = i\}, \quad i \in S \]

- Required to estimate both \( x_k \) (continuous-valued) and \( r_k \) (discrete-valued):
  Hybrid estimation!
Error Bounds for switching dynamic models

- Impossible to derive exact Cramer-Rao lower bounds
  
  Requires differentiation of terms such as $\log p(r_{k+1}|r_k)$

- Alternatives:
  
  1. Explore more general bounds than the Cramer-Rao bound
     
     e.g. Bhattacharya, Bobovsky-Zakai, Weiss-Weinstein lower bounds
      
     Problem: computationally expensive!
  
  2. Develop an approximate Cramer-Rao lower bound
     
     a. Conditioning on the regime sequence (i.e. enumeration bound)
      
     b. Using best fitting Gaussian distributions [Hernandez, Ristic, Farina, 2005]
Conditioning on the regime sequence (enumeration bound)

- Let: $\rho_{nk}^n \triangleq (r_{n1}^n, \ldots, r_{nk}^n)$ be $n$-th regime sequence ($n = 1, 2, \ldots, s^k$)

- Then easily shown:

\[
\mathbb{E}\left\{ (\hat{x}_k - x_k) [\hat{x}_k - x_k]^T \right\} \geq \sum_{n=1}^{s^k} \mathbb{P}(\rho_{nk}^n) \cdot [J_{nk}^n]^{-1}
\]

- The RHS gives the *enumeration* bound

- $\mathbb{P}(\rho_{nk}^n)$ is the (prior) probability of sequence $\rho_{nk}^n$; can be computed knowing initial $p_1(i)$ and transitional $\pi_{ij}$ regime probabilities.

- $J_{nk}^n$ is the (Fisher) information matrix conditional on sequence $\rho_{nk}^n$:

\[
J_{nk}^n = \left[ \Sigma_{k-1}(r_{nk}^n) + F_{k-1}(r_{nk}^n) [J_{k-1}^n]^{-1} F_{k-1}(r_{nk}^n)^T \right]^{-1} + J_z(k)
\]
**Conditioning on \( \rho^n_k \): Optimistic bound**

- Each \( [J^n_k]^{-1} \) gives the error covariance bound for a known manoeuvre sequence \( \rho^n_k \);
  \( \Rightarrow \) the resulting CR bound is overly optimistic!

- Demonstration of this over-optimism with simple example:
  - \( S1 \): target in either CT or NCV model (manoeuvring)
  - \( S2 \): target always in NCV model
  - measurements linear in target state in both cases: hence \( J_z(k) \) same

- We expect the CRB for \( S1 \) (manoeuvring target) to be higher as a consequence of additional uncertainty due to model switching.
Conditioning on $\rho_{nk}^n$: Optimistic bound demonstration
Switching Dynamic models: Best fitting Gaussian

- Original model (MODEL 1):
  \[ x_{k+1} = F_k(r_{k+1})x_k + w_k(r_{k+1}) \text{ with } w_k(r_{k+1}) \sim \mathcal{N}(0, \Sigma_k(r_{k+1})) \]

- Replace with a best-fitting Gaussian (BFG) approximation (MODEL 2):
  \[ x_{k+1} \approx \Phi_k x_k + \epsilon_k \text{ with } \epsilon_k \sim \mathcal{N}(0, Q_k) \]

- \( \Phi_k \) and \( Q_k \) chosen so that:
  \[ \mathbb{E}[x_k | \text{MODEL 1}] = \mathbb{E}[x_k | \text{MODEL 2}] \text{ for all } k \]

  \[ Cov[x_k | \text{MODEL 1}] = Cov[x_k | \text{MODEL 2}] \text{ for all } k \]

- \( Q_k \) must also be positive definite (being a covariance)
Switching Dynamic models: Best fitting Gaussian (Cont’d)

- The BFG-CRB is then simply computed using the Riccati-like recursion:

\[
J_{k+1} = \left( Q_k + \Phi_k J^{-1}_k \Phi^T_k \right)^{-1} + J_z(k + 1)
\]

- Initialisation:

  - Assuming that the prior pdf is: \( x_0 \sim N(\bar{x}_0, P_0) \), set:

    \[
    \varepsilon_0 = \bar{x}_0, \quad C_0 = P_0
    \]

  - Determine mode probabilities:

    - **define:** \( p_k(r) \triangleq \mathbb{P}(r_k = r) \), for \( r = 1, \ldots, s \)

    - **determine:** \( p_k(r) = \sum_{j=1}^{s} \pi_{jr} p_{k-1}(j) \) for \( k = 2, 3 \ldots \)
**BFG Distribution – General Recursion**

- **STEP 1:** determine $\Phi_k$ as follows:
  \[
  \Phi_k = \sum_{r=1}^{s} F_k(r) p_{k+1}(r)
  \]

- **STEP 2:** determine $C_{k+1}$ as follows:
  \[
  C_{k+1} = \sum_{r=1}^{s} p_{k+1}(r) \left[ F_k(r) \left( C_k + \varepsilon_k \varepsilon_k^T \right) F_k^T(r) + \Sigma_k(r) \right] - \Phi_k \varepsilon_k \varepsilon_k^T \Phi_k^T
  \]

- **STEP 3:** determine $Q_k$ as follows:
  \[
  Q_k = C_{k+1} - \Phi_k C_k \Phi_k^T
  \]
  (guaranteed $Q_k \geq 0$)

- **STEP 4:** determine $\varepsilon_{k+1}$ as follows:
  \[
  \varepsilon_{k+1} = \Phi_k \varepsilon_k
  \]

- **STEP 5:** set: $k \rightarrow (k + 1)$ and repeat from **STEP 1**
BFG CR Bound demonstration

BFG approximation incorporates uncertainty due to model switching
**Verification of the BFG approximation**

- **Aim:** Compare the theoretical bound with empirical RMS error performance
- **We simulate a target switching between CV or CA models (no process noise);**
- **Transition probabilities:** $\pi_{ii} = 0.9$ for $i = 1, 2$
- **Sampling time** $T = 3$ seconds
- **Measurements of Cartesian coordinates; error standard deviations:** $\sigma_x = \sigma_y = 200$ m
- **Comparison between:**
  - Two theoretical CR bounds (BFG bound and Enumeration bound)
  - Empirical RMS error of an IMM filter; obtained via Monte Carlo simulations.
Verification of the BFG approximation (Cont’d)
The effect of $P_d < 1$ and $P_{fa} > 0$

- Most sensors characterised by $P_d < 1$ and $P_{fa} > 0$  
  ⇒ Uncertainty in measurement origin

- This type of uncertainty affects only $J_z(k)$ in: $J_k = J_p(k) + J_z(k)$

- Several contributions since 1990 (more than 10 publications, Jauffret, Bar-Shalom, Zhang, Willet, Hernandez, Farina, Ristic, etc)

- The most comprehensive treatment (captures all previous developments) is the measurement sequence conditioning approach:

Measurement sequence conditioning

- Measurements sequence: \( M_{1:k} = \{m_1, m_2, \ldots, m_k\} \)

- \( m_i \) is the number of measurements received at time \( i = 1, \ldots, k \).
  \( m_i \in \{0, 1, 2, \ldots\} \)

- The CR inequality is then:
  \[
  \mathbb{E} \left\{ (\hat{x} - x) (\hat{x} - x)^T \right\} \geq \sum_{M_{1:k}} P(M_{1:k}) J_k^{-1}(M_{1:k})
  \]

- \( P(M_{1:k}) \) can be computed knowing:
  - the probability of detection \( P_d \)
  - the expected number of false measurements in the gate (Poisson model)
Measurement sequence conditioning (Cont’d)

- Information matrix as always have two components:

\[
J_k(M_{1:k}) = J_p(k : M_{1:k-1}) + J_z(k : m_k)
\]

- Under some reasonable assumptions (rectangular gates, diagonal measurement matrix \(R_k\)) we obtain:

\[
J_z(k : m_k) = q_k(m_k) \mathbb{E}\{H_k^T R_k^{-1} H_k\}
\]

where \(q_k(m_k)\) is the information reduction factor (needs to be computed numerically);

- If \(P_d = 1\) and \(P_{fa} = 0\), then \(m_k = 1\) and \(q_k(1) = 1\) (see slide 10).
Measurement sequence conditioning: No false alarms

- \( m_k \in \{0, 1\} \)

- Sequence \( M_{1:k} \) becomes a “detection/miss” sequence, so that:

  \[
  J_z(k : m_k) = \begin{cases} 
  0 & \text{if } m_k = 0, \\
  \mathbb{E}\{H_k^T R_k^{-1} H_k\} & \text{if } m_k = 1.
  \end{cases}
  \]


- When the false alarm rate is small (e.g. average number of false detections in the gate is below 0.1), the CR bound mainly influenced by \( P_d < 1 \).
The influence of $P_d < 1$

Tracking a ballistic object on re-entry (slide 20)

Multiple target tracking

- Notoriously difficult if multiple targets appear and disappear at random: the problem requires joint detection and tracking; Cramér-Rao bound not a suitable tool!

- If we assume that $L \geq 1$ targets exist in a surveillance region during the observation period, possible to formulate a CRB: Hue et al. [IEEE AES 2006], Tharmarasa et al [IEEE AES 2006].

- An analytic expression for multi-target CR bound in the framework of track-before-detect (ultimate bound)
  - Depends on SNR, sensor resolution, point-spread function and target kinematics.
  - Directly applicable to Wireless Sensor Networks (WSN)
Example: Wireless network of acoustic sensors

- State vector: \( \mathbf{x}_{k,i} = [x_{k,i}, \dot{x}_{k,i}, y_{k,i}, \dot{y}_{k,i}, A_{k,i}]^T; \)
  \( i = 1, 2, \ldots \) is target index

- Target motion nearly CV

- Location of sensor \( j \) is: \( (X^j, Y^j), j = 1, 2, \ldots, N_s \)

- Measurements of sound intensity (at sensor \( j \)):

\[
\tilde{z}^j_k = \sum_i \frac{A_{k,i}}{\sqrt{(X^j - x_{k,i})^2 + (Y^j - y_{k,i})^2}} + v^j_k
\]
Example: Wireless network of acoustic sensors (Cont’d)

- The (sound) intensity of the blue target is 3 dB higher

- Easy to include the effects of quantisation, and to predict the required sensor density.
A Sensor Management Application

- Context:
  - Tracking of an anti-ship missile using a combination of a phased-array radar and an IRST sensor.
  - The IRST passively scans the horizon at a constant scanning interval in order to detect low altitude threats; each detection serves as an alert to allocate and cue the radar.

- The Cramér-Rao bound analysis applied to predict an average radar allocation requirements as a function of: target manoeuvrability, sensor accuracy, positional estimation accuracy.
Average radar update time

Versus (a) IRST sampling interval; (b) missile manoeuvrability
Summary

- Cramér-Rao bounds enable us to quantify the (best achievable) tracking error performance;

- A useful tool for tracker design, algorithm assessment, sensor management, etc.

- Significant progress made in the last few years on the CRB development for tracking

- Shortcomings:
  - impossible to compute in all situations (e.g. appearance of targets, switching models)
  - in some cases cannot be achieved by any practical estimator
Future work

- Multi-target tracking, hard constraints, comparison of *ultimate* bound with the *thresholding* bound, etc.

- Explore other variance bounds
  (Bhattacharya, Bobovsky-Zakai, Weiss-Weinstein, Barankin, etc)