Detailed Statistical Analysis for the Proposed Estimator

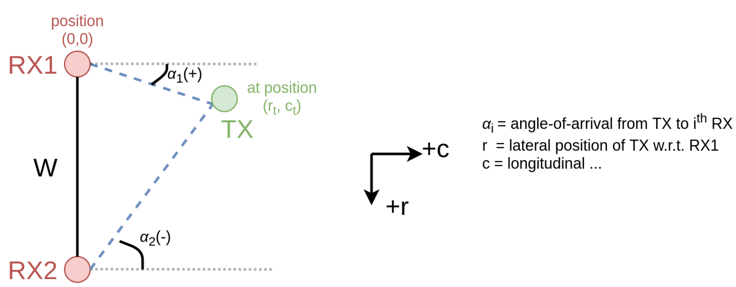
Burak Soner, Sinem Coleri

We're interested in analyzing the performance of a specific unbiased 2D location estimation algorithm (the one proposed in the SonerColeri TVT article) that uses noisy angle measurements as input. In order to do this:

1. we first define the observation model for this problem and present the estimation algorithm that uses those noisy observations (angle-of-arrival measurements) for position estimation.
2. we derive Cramer-Rao Lower Bound (CRLB) for the observation model and show that the estimator approximately attains this bound during simulations for configurations that are of interest to the application.
3. however, although the bound is approximately attained in simulation for nearly all realistic scenarios, we observed that the estimator does not satisfy the CRLB theorem (we'll show this in detail below). This means that the estimator is not efficient, but it could still be the minimum variance unbiased (MVU) estimator, i.e., the best possible unbiased estimator for this problem in the mean-squared error sense.
4. we next apply the Rao-Blackwell-Lehmann–Scheffé theorem (RBLS, considers sufficient and complete statistics) to this problem to prove that the estimator actually is the MVU estimator for this problem.
5. furthermore, we also show that this estimator is the maximum likelihood estimator (MLE) for this problem.

1. System Description

Consider the following wireless communication system with 2 transmitters (TX) and 2 receivers (RX):



The RX units are at fixed locations with a baseline of W between them, RX1 is located at the origin of the coordinate system, RX2 does not move, and the TX position, i.e., (r_t, c_t) is measured with respect to the origin. The TX can move freely. Both RX units can measure the angle-of-arrival (AoA) from TX signals based on the characteristics of received signals. The AoA measurements are then used for estimating relative TX position based on the following relations (uses the sine rule) (this is exactly the estimator in the TVT article):

$$\widehat{r_t} = W\left(1 + \frac{\sin(\widehat{\alpha_2}) \;\cdot\; \cos(\widehat{\alpha_1})}{\sin(\widehat{\alpha_1} - \widehat{\alpha_2})}\right) \quad , \quad \widehat{c_t} = W\left(\frac{\cos(\widehat{\alpha_2}) \;\cdot\; \cos(\widehat{\alpha_1})}{\sin(\widehat{\alpha_1} - \widehat{\alpha_2})}\right) \tag{1}$$

where $(\widehat{r_t}, \widehat{c_t})$ are estimations for the relative TX coordinates, and $(\widehat{\alpha_1}, \widehat{\alpha_2})$ are the observations (i.e., the AoA measurements) which are contaminated due to channel noise. Note that this is a vector estimation problem (i.e., observation vector size = 2x1, parameter vector size = 2x1), and we only take 1 sample observation vector (of size 2x1) to produce 1 estimation for the parameter vector (of size 2x1). We assume W is exactly known. The AoA measurement model, i.e., the observation model, is as follows:

$$\widehat{\alpha_1} = \alpha_1 + \eta_1 = \arctan\left(\frac{r_t}{c_t}\right) + \eta_1 \quad , \quad \widehat{\alpha_2} = \alpha_2 + \eta_2 = \arctan\left(\frac{r_t}{c_t}\right) + \eta_2$$

$$\widehat{\mathbf{x}} = [\widehat{\alpha_1} \;\; \widehat{\alpha_2}] \quad , \quad \mathbf{x} = [\alpha_1 \;\; \alpha_2]$$

where η_1 and η_2 are independent additive white Gaussian noise (AWGN) components for each AoA measurement with variance σ_η^2, $\widehat{\mathbf{x}}$ is the observation vector, and $\mathbf{x}$ is the deterministic ("true") AoA vector. This means $\widehat{\alpha_1}$ and $\widehat{\alpha_2}$ are Gaussian random variables with mean values equal to the true AoA values (α_1, α_2) and variances equal to σ_η^2. Since the expected AoA values produce the true location when inserted into Eqn. (1), the estimator is unbiased.

2. Estimator gets very close to CRLB in simulation

We derive the CRLB for this problem by finding elements of the Fisher Information Matrix (FIM). We could always take the 2nd derivatives of the log-likelihood fcn here and compute the (negative) expected values to build the FIM the conventional way, but there's a much easier way to compute those elements in our case using (3.33) from Kay:

"*Generalizing to a vector signal parameter estimated in the presence of WGN,* [we have, in our terms]*:*

$$[\mathbf{I}(\theta)]_{m,m'} = \frac{1}{\sigma_\eta^2} \sum_{n=1}^{4} \left(\frac{\partial \mathbf{x}[n]}{\partial \theta_{\mathbf{m}}} \cdot \frac{\partial \mathbf{x}[n]}{\partial \theta_{\mathbf{m'}}} \right) \tag{2}$$

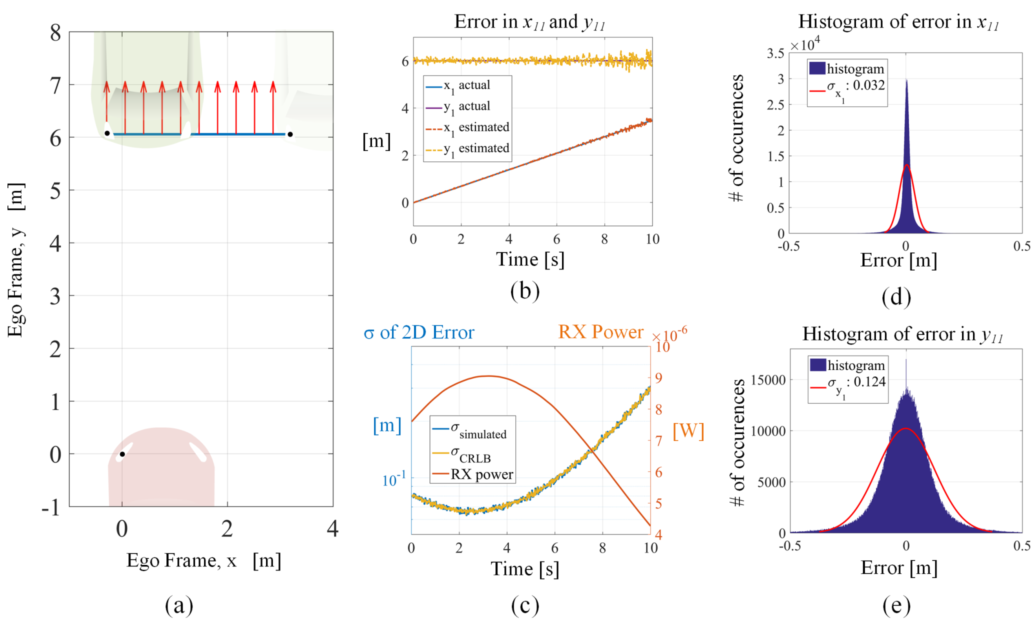
as the elements of the Fisher information matrix. "

OK let's do that using Eqn. (2):

$$\mathbf{I_{11}} = \frac{1}{\sigma_\eta^2} \cdot \left[\left(\frac{c_{11}}{r_{11}^2 + c_{11}^2} \right)^2 + \left(\frac{c_{11}}{(r_{11} - W)^2 + c_{11}^2} \right)^2 \right] \quad , \quad \mathbf{I_{12}} = -\frac{1}{\sigma_\eta^2} \cdot \left[\frac{r_{11} \cdot c_{11}}{(r_{11}^2 + c_{11}^2)^2} + \frac{r_{11} \cdot c_{11}}{\left((r_{11} - W)^2 + c_{11}^2 \right)^2} \right]$$

$$\mathbf{I_{22}} = \frac{1}{\sigma_\eta^2} \cdot \left[\left(\frac{-r_{11}}{r_{11}^2 + c_{11}^2} \right)^2 + \left(\frac{-(r_{11} - W)}{(r_{11} - W)^2 + c_{11}^2} \right)^2 \right] \quad , \quad \mathbf{I_{21}} = \mathbf{I_{12}}$$

The diagonal elt.s of the inverse of the FIM provide the lower bound on the variance any unbiased estimator can attain for this problem. Let's evaluate this for certain configurations [1], and see if our estimator gets close to this:



This figure shows the performance of the estimator in Eqn. (1) and its CRLB over a certain vehicular trajectory (the TX is the left tail-light of the green target vehicle and RXs are detectors mounted on the head-lights of the red ego vehicle). The errors in AoA measurement are zero-mean AWGN with variance σ_η^2 here [2], and in (c) we see a very small difference between the CRLB and the actual variance of the estimator output, i.e., **the estimator seems like it really does attain the CRLB**.

[1] we need to sample the variance of the estimator output through multiple trials. We repeat the test 1000 times to do this

[2] just a note: as the histograms in (d) and (e) show, the propagation of the Gaussian AoA errors does not result in a Gaussian distribution of location estimate errors (i.e., x, y errors in the figure). This is due to the non-linearity of the system model from AoA $\rightarrow$ location parameters. In very high signal-to-noise-ratio (SNR) cases, the shape of the location parameter error distributions do get closer to Gaussian curves, but this is not specifically a low-SNR scenario considering realistic channel conditions, and we still see the effect.

3. Estimator gets close to the CRLB, but it's not efficient, might still be the MVU

We try using the CRLB equality condition to assess whether or not our estimator is statistically efficient; the CRLB equality condition is defined as follows in (3.7) from Kay:

"*An unbiased estimator may be found that attains the bound for all θ if and only if*

$$\frac{\partial \ln p(\widehat{\mathbf{x}};\theta)}{\partial \theta} = \mathbf{I}(\theta)\left(\mathbf{g}(\widehat{\mathbf{x}}) - \theta\right) \tag{3}$$

for some functions $\mathbf{g}$ *and* $\mathbf{I}$*. That estimator, which is the minimum variance unbiased estimator, is* $\widehat{\theta} = \mathbf{g}(\mathbf{x})$*, and the minimum variance* [for each parameter is contained in the diagonal elements of $\mathbf{I^{-1}}(\theta)$] ".

The function $\mathbf{I}$ here is the FIM, which we already computed, and we already know the estimator g, so we only need the left hand side. Let's start from the likelihood function (i.e., the PDF) for our system:

$$p(\widehat{\mathbf{x}};\theta) = \frac{1}{\left(2\pi\sigma_\eta^2\right)^{\left(\frac{2}{2}\right)}} \cdot \exp\left(-\frac{1}{2\sigma_\eta^2} \cdot \left(\left(\widehat{\alpha_1} - \arctan\left(\frac{r_t}{c_t}\right)\right)^2 + \left(\widehat{\alpha_2} - \arctan\left(\frac{r_t - W}{c_t}\right)\right)^2\right)\right)$$

and its log version:

$$\ln p(\widehat{\mathbf{x}};\theta) = -\ln 2\pi\sigma_\eta^2 - \frac{1}{2\sigma_\eta^2} \cdot \left(\left(\widehat{\alpha_1} - \arctan\left(\frac{r_t}{c_t}\right)\right)^2 + \left(\widehat{\alpha_2} - \arctan\left(\frac{r_t - W}{c_t}\right)\right)^2\right)$$

The first derivatives of the log-likelihood function with respect to the parameter vector elements:

$$\begin{aligned}
\frac{\partial \ln p(\widehat{\mathbf{x}};\theta)}{\partial r_t} &= -\frac{1}{\sigma_\eta^2} \cdot \left[\left(\widehat{\alpha_1} - \arctan\left(\frac{r_t}{c_t}\right)\right)\left(-\frac{\partial \arctan\left(\frac{r_t}{c_t}\right)}{\partial r_t}\right) + \left(\widehat{\alpha_2} - \arctan\left(\frac{r_t - W}{c_t}\right)\right)\left(-\frac{\partial \arctan\left(\frac{r_t - W}{c_t}\right)}{\partial r_t}\right)\right] \\
&= -\frac{1}{\sigma_\eta^2} \cdot \left[\left(\widehat{\alpha_1} - \arctan\left(\frac{r_t}{c_t}\right)\right)\left(-\frac{c_t}{r_t^2 + c_t^2}\right) + \left(\widehat{\alpha_2} - \arctan\left(\frac{r_t - W}{c_t}\right)\right)\left(-\frac{c_t}{\left(r_t - W\right)^2 + c_t^2}\right)\right] \\
&= \frac{1}{\sigma_\eta^2} \cdot \left[\left(\widehat{\alpha_1} - \arctan\left(\frac{r_t}{c_t}\right)\right)\left(\frac{c_t}{r_t^2 + c_t^2}\right) + \left(\widehat{\alpha_2} - \arctan\left(\frac{r_t - W}{c_t}\right)\right)\left(\frac{c_t}{\left(r_t - W\right)^2 + c_t^2}\right)\right] \qquad (4)
\end{aligned}$$

(skipping the simplification steps in the next term):

$$\frac{\partial \ln p(\widehat{\mathbf{x}};\theta)}{\partial c_t} = -\frac{1}{\sigma_\eta^2} \cdot \left[\left(\widehat{\alpha_1} - \arctan\left(\frac{r_t}{c_t}\right)\right)\left(\frac{r_t}{r_t^2 + c_t^2}\right) + \left(\widehat{\alpha_2} - \arctan\left(\frac{r_t - W}{c_t}\right)\right)\left(\frac{r_t - W}{\left(r_t - W\right)^2 + c_t^2}\right)\right] \tag{5}$$

Now we have all the necessary terms for the left hand side (LHS) and the right hand (RHS) side of Eqn. (3). However, the terms are a bit ugly, so let's try to see if numerically evaluating the two sides gives the same result.

Simply simulate with numpy for an example configuration:

```
def xy_from_psi12_W(theta1, theta2, W):
    return (W*(1+(np.sin(theta2)*np.cos(theta1) / np.sin(theta1-theta2))),
            W*(np.cos(theta2)*np.cos(theta1) / np.sin(theta1-theta2))      )
```

```
# we tilt the TX "stick" a bit. r12 and c12 are computed accordingly, based on r11, c11 and tilt
target_tilt = np.pi/32;

W   = 1
r11 = 1.2
c11 = 2.5
r12 = r11 + W*np.cos(target_tilt)
c12 = c11 + W*np.sin(target_tilt)
print("tx1 pos:", (r11, c11), "tx2 pos:",(r12, c12))

a11_true = np.arctan2(r11, c11)
a12_true = np.arctan2(r12, c12)
a21_true = np.arctan2(r11 - W, c11)
a22_true = np.arctan2(r12 - W, c12)
```

```
tx1 pos: (1.2, 2.5) tx2 pos: (2.195184726672197, 2.5980171403295604)
```

```
noise_mean  = 0
noise_stdev = 0.001 # in radians!
noise1      = np.random.normal(noise_mean, noise_stdev, 1) # iid
noise2      = np.random.normal(noise_mean, noise_stdev, 1) # iid
noise3      = np.random.normal(noise_mean, noise_stdev, 1) # iid
noise4      = np.random.normal(noise_mean, noise_stdev, 1) # iid

a11_noised = a11_true + noise1
a12_noised = a12_true + noise2
a21_noised = a21_true + noise3
a22_noised = a22_true + noise4

r11_est, c11_est = xy_from_psi12_W(a11_noised, a21_noised, W)
r12_est, c12_est = xy_from_psi12_W(a12_noised, a22_noised, W)
print("tx1 est. pos:", (r11_est.item(), c11_est.item()), "tx2 est. pos:", (r12_est.item(), c12_est.item()))
```

```
tx1 est. pos: (1.1996718935797726, 2.496810198003803) tx2 est. pos: (2.1869445299547032, 2.5784518005465578)
```

evaluate left hand side, i.e, first derivatives

```
dlnp_dr11 = (1/(noise_stdev**2))*( (a11_noised - np.arctan(r11/c11))*(c11/(r11**2+c11**2)) + (a21_noised - np.arctan(
dlnp_dc11 =-(1/(noise_stdev**2))*( (a11_noised - np.arctan(r11/c11))*(r11/(r11**2+c11**2)) + (a21_noised - np.arctan(
dlnp_dr12 = (1/(noise_stdev**2))*( (a12_noised - np.arctan(r12/c12))*(c12/(r12**2+c12**2)) + (a22_noised - np.arctan(
dlnp_dc12 =-(1/(noise_stdev**2))*( (a12_noised - np.arctan(r12/c12))*(r12/(r12**2+c12**2)) + (a22_noised - np.arctan(
print(dlnp_dr11, dlnp_dc11, dlnp_dr12, dlnp_dc12)
```

```
[115.74711357] [-60.17236738] [498.01317392] [-391.02870548]
```

evaluate right hand side, the one with the FIM

```
i11 =  (1/(noise_stdev**2))*( (c11/(r11**2 + c11**2))**2 + (c11/((r11-W)**2 + c11**2))**2 )
i12 = -(1/(noise_stdev**2))*( (r11*c11/((r11**2 + c11**2)**2)) + ((c11*(r11-W))/(((r11-W)**2 + c11**2)**2)) )
i21 =  i12
i22 =  (1/(noise_stdev**2))*( (-r11/(r11**2 + c11**2))**2 + (-(r11-W)/((r11-W)**2 + c11**2))**2 )

i33 =  (1/(noise_stdev**2))*( (c12/(r12**2 + c12**2))**2 + (c12/((r12-W)**2 + c12**2))**2 )
i34 = -(1/(noise_stdev**2))*( (r12*c12/((r12**2 + c12**2)**2)) + ((c12*(r12-W))/(((r12-W)**2 + c12**2)**2)) )
i43 =  i34
i44 =  (1/(noise_stdev**2))*( (-r12/(r12**2 + c12**2))**2 + (-(r12-W)/((r12-W)**2 + c12**2))**2 )

rhs_r11 = i11*( r11_est - r11 ) + i12*( c11_est - c11 )
rhs_c11 = i21*( r11_est - r11 ) + i22*( c11_est - c11 )
rhs_r12 = i33*( r12_est - r12 ) + i34*( c12_est - c12 )
rhs_c12 = i43*( r12_est - r12 ) + i44*( c12_est - c12 )
print(rhs_r11, rhs_c11, rhs_r12, rhs_c12)
```

```
[115.62335646] [-60.10706649] [494.93366329] [-388.64753672]
```

(image is from the formulation with two TXs but just check the first two elements for each 4x1 vector)

The LHS and RHS are pretty close, but they're not exactly the same, i.e., there is a finite difference e=LHS-RHS.

The CRLB theorem states the efficiency condition with "if and only if" and our estimator does not meet this.

So although the simulation results and the CRLB are very very close, our estimator is not efficient. It might still be the MVU though, this analysis doesn't prove/disprove anything about that.

4. Prove that the estimator is the MVU via RBLS theorem

To prove that our estimator is the MVU for this problem, we need to use the RBLS theorem, which states:

Theorem 5.4 (Rao-Blackwell-Lehmann-Scheffe (Vector Parameter)) *If $\check{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$ and $\mathbf{T}(\boldsymbol{\theta})$ is an $r \times 1$ sufficient statistic for $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\theta}} = E(\check{\boldsymbol{\theta}}|\mathbf{T}(\mathbf{x}))$ is*

1. *a valid estimator for $\boldsymbol{\theta}$ (not dependent on $\boldsymbol{\theta}$)*
2. *unbiased*
3. *of lesser or equal variance than that of $\check{\boldsymbol{\theta}}$ (each element of $\hat{\boldsymbol{\theta}}$ has lesser or equal variance)*

Additionally, if the sufficient statistic is complete, then $\hat{\boldsymbol{\theta}}$ is the MVU estimator.

So we first need to find a sufficient statistic for our problem, show that it is complete, and then show that our estimator (which we already know to be unbiased) actually uses that statistic (which we'll show here to be sufficient and complete) as its input. If we succeed in this, our estimator is guaranteed to be the MVU estimator for this problem.

4.1 Finding a sufficient statistic for this problem

Usually, we need to use Neyman-Fisher Factorization theorem to find sufficient statistics for a given problem, but in our case the solution is actually obvious → we already use all the data that we have available as the input of the estimator, i.e., the single sample observation vector $\widehat{x} = [\widehat{\alpha_1}, \widehat{\alpha_2}]$, and the dataset itself is always a sufficient statistic for the dataset. Kay also states this while explaining the Neyman-Fisher Factorization theorem:

Note that there always exists a set of sufficient statistics, the data set $\mathbf{x}$ itself being sufficient.

Therefore, **our sufficient statistic is the data itself**.

4.2 Showing that our sufficient statistic is complete

Since the elements of our single-sample-observation-vector (of size 2x1) are i.i.d. Gaussian random variables (i.e., $\widehat{\alpha_1}$ and $\widehat{\alpha_2}$), we can directly employ the proof for completeness provided in Example 5.6 from Kay, which considers the sample mean estimator (observations are samples from Gaussian r.v. with mean A, the true value, so the unbiased estimator is the sample mean), and the sum of all observations as the sufficient statistic:

Example 5.6 - Completeness of a Sufficient Statistic

For the estimation of A, the sufficient statistic $\sum_{n=0}^{N-1} x[n]$ is complete or there is but one function g for which $E[g(\sum_{n=0}^{N-1} x[n])] = A$. Suppose, however, that there exists a second function h for which $E[h(\sum_{n=0}^{N-1} x[n])] = A$. Then, it would follow that with $T = \sum_{n=0}^{N-1} x[n]$,

$$E\left[g(T) - h(T)\right] = A - A = 0 \qquad \text{for all } A$$

or since $T \sim \mathcal{N}(NA, N\sigma^2)$

$$\int_{-\infty}^{\infty} v(T) \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{1}{2N\sigma^2}(T - NA)^2\right] dT = 0 \qquad \text{for all } A$$

where $v(T) = g(T) - h(T)$.

Letting $\tau = T/N$ and $v'(\tau) = v(N\tau)$, we have

$$\int_{-\infty}^{\infty} v'(\tau) \frac{N}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{N}{2\sigma^2}(A-\tau)^2\right] d\tau = 0 \qquad \text{for all } A \tag{5.7}$$

which may be recognized as the convolution of a function $v'(\tau)$ with a Gaussian pulse $w(\tau)$ (see Figure 5.3). For the result to be zero for all A, $v'(\tau)$ must be identically zero. To see this recall that a signal is zero if and only if its Fourier transform is identically zero, resulting in the condition

$$V'(f)W(f) = 0 \qquad \text{for all } f$$

where $V'(f) = \mathcal{F}\{v'(\tau)\}$ and $W(f)$ is the Fourier transform of the Gaussian pulse in (5.7). Since $W(f)$ is also Gaussian and therefore positive for all f, we have that the condition is satisfied if and only if $V'(f) = 0$ for all f. Hence, we must have that $v'(\tau) = 0$ for all τ. This implies that $g = h$ or that the function g is unique. $\diamond$

Although we have the following differences from this example:

- we employ the dataset itself, rather than the sum of all values, as the sufficient statistic,
- our system model is non-linear, rather than the linear identity function in the example,
- we have a single sample of a vector observation, rather than N samples for a scalar observation,

these differences do not change the result since the completeness properties for the Gaussian PDF also hold in the vector case. Therefore, **our sufficient statistic is also complete**, by the same proof.

4.3 Showing that we our estimator is unbiased using the above statistic as its input

This one is trivial since we already formulated our estimator such that it is unbiased and it uses the single-sample-observation-vector $\widehat{x} = [\widehat{\alpha_1}, \widehat{\alpha_2}]$, which was shown to be a sufficient and complete statistic for this problem.

Having satisfied all conditions for the RBLS theorem, we now draw the following conclusions (including ones we arrived at in previous subsections):

- The variance of our unbiased estimator gets really close to the CRLB in simulation,
- but it is not an efficient estimator because it does not satisfy the CRLB theorem.
- However, through the RBLS theorem we see that it is the MVU estimator for this problem.

5. Show that the estimator is also the MLE for this problem

We do this by setting the first derivatives of the log-likelihood function that we derived in Section 3, in Eqns. (4) and (5), to 0, and check whether this is satisfied when our estimator is plugged in. In other words, we plug in the estimator expressions in Eqn. (1) to Eqns. (4) and (5) and see if the two equations can analytically be shown to be =0.

We can easily do this by showing that $\widehat{r_t}/\widehat{c_t}$ becomes $\tan(\widehat{\alpha_1})$ and $(\widehat{r_t} - W)/\widehat{c_t}$ becomes $\tan(\widehat{\alpha_2})$, which makes the two terms in addition in Eqns. (4) and (5) equal to 0:

$$\frac{\widehat{r_t}}{\widehat{c_t}} = \frac{W \cdot \left(1 + \frac{\sin(\alpha_2)\cdot\cos(\alpha_1)}{\sin(\alpha_1-\alpha_2)}\right)}{W \cdot \left(\frac{\cos(\alpha_2)\cdot\cos(\alpha_1)}{\sin(\alpha_1-\alpha_2)}\right)} = \frac{\sin(\alpha_1-\alpha_2) + \sin(\alpha_2)\cdot\cos(\alpha_1)}{\cos(\alpha_2)\cdot\cos(\alpha_1)} =$$

$$\frac{\sin(\alpha_1)\cdot\cos(\alpha_2) - \sin(\alpha_2)\cdot\cos(\alpha_1)}{\cos(\alpha_2)\cdot\cos(\alpha_1)} + \tan(\alpha_2) \quad = \quad \tan(\alpha_1) + \tan(\alpha_2) - \tan(\alpha_2) \quad = \quad \tan(\alpha_1) \tag{6}$$

showing this holds for $(\widehat{r_t} - W)/\widehat{c_t} = \tan(\alpha_2)$ is straightforward. Since this analytically shows that our estimator makes Eqns. (4) and (5) = 0, **we conclude that our estimator is also the MLE**.