Bias-Resolution-Variance Tradeoffs for Single Pixel Estimation Tasks using the Uniform Cramér-Rao Bound

Thomas Kragh
EECS Department
University of Michigan
Ann Arbor, MI

Outline

- Problem Statement
- Unbiased Estimator Cramér-Rao (CR) Bound
- Biased Estimator CR Bound
- Uniform CR Bound (UCRB)
- Interpretations of the UCRB
- Problems with UCRB
- 2nd-moment as measure of resolution
- UCRB with resolution constraint
- Pretty Pictures

Problem Statement

Consider estimation of a vector parameter $\underline{\theta} = [\theta_1, \ldots, \theta_n]^T$,
given an observation of the vector random variable $\underline{Y}$
with probability density function $f_{\underline{Y}}\left(\underline{y} | \underline{\theta}\right)$.

Specifically, consider the biased single-component estimator
$\hat{\theta}_p = \underline{e}_p{}^T \hat{\underline{\theta}}(\underline{Y})$, where $\underline{e}_p = [0, \ldots, 1, \ldots, 0]^T$ (p$^{\text{th}}$–unit basis vector)

Derive an expression for a lower bound of the estimator variance.

Unbiased Estimator
Cramér-Rao (CR) Bound

$$\operatorname{var}_{\underline{\theta}}\left(\hat{\theta}_p\right) \geq \underline{e}_p{}^T F_{\underline{Y}}^{-1} \underline{e}_p = \left[F_{\underline{Y}}^{-1}\right]_{pp}$$

- Fisher Information $F_{\underline{Y}} = E_{\underline{\theta}}\left[\nabla \ln f_{\underline{Y}}\left(\underline{y}|\underline{\theta}\right) \cdot \nabla^T \ln f_{\underline{Y}}\left(\underline{y}|\underline{\theta}\right)\right]$

- Gradient operator $\nabla = \left[\frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_n}\right]^T$

Example: 1D estimation

- Measure $y = \theta + n$
  - θ is an (unknown) constant
  - *n* is zero-mean gaussian random variable with variance σ^2

- From a *single* observation Y of the random variable y, how well can we estimate θ?

- Fisher Information: $F_Y(Y) = \frac{1}{\sigma^2}$

- CR-Bound (unbiased estimators) : $\mathrm{var}(\hat{\theta}) \geq F_Y^{-1} = \sigma^2$

Derive the ML-estimator

- For a fixed measurement value Y, define the likelihood function *L(θ)* as

$$L(\theta) = f_{y|\theta}(Y|\theta)$$

- The ML-Estimator is defined as

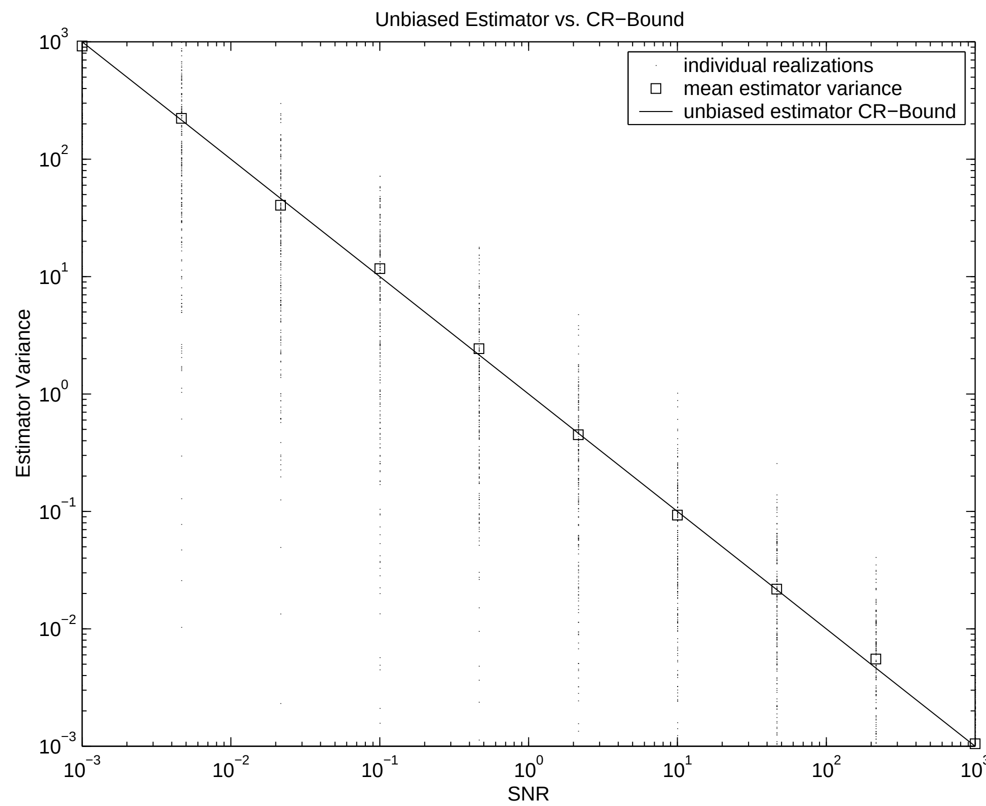
$$\hat{\theta}_{ML}(Y) = \underset{\theta}{\text{arg - max}} \; L(\theta)$$

- For our case,

$$f_{y|\theta}(Y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y-\theta)^2}{2\sigma^2}} \quad \Rightarrow \quad \boxed{\hat{\theta}_{ML}(Y) = Y}$$

Numerical Simulation of ML-Estimator

- ML-Estimator $\hat{\theta} = Y$
- Estimator variance measured by sample mean
- 200 realizations at each SNR
- $SNR = \dfrac{\theta^2}{\sigma^2}$
- $\theta = 1$



Try something new:
Include some apriori knowledge...

- Let $y = \theta + n$ as before
- Assume we know apriori that θ > 0
- ML-Estimator is therefore …

$$\hat{\theta}_{ML}(Y) = \underset{\theta \geq 0}{\text{arg - max}} \; L(\theta)$$

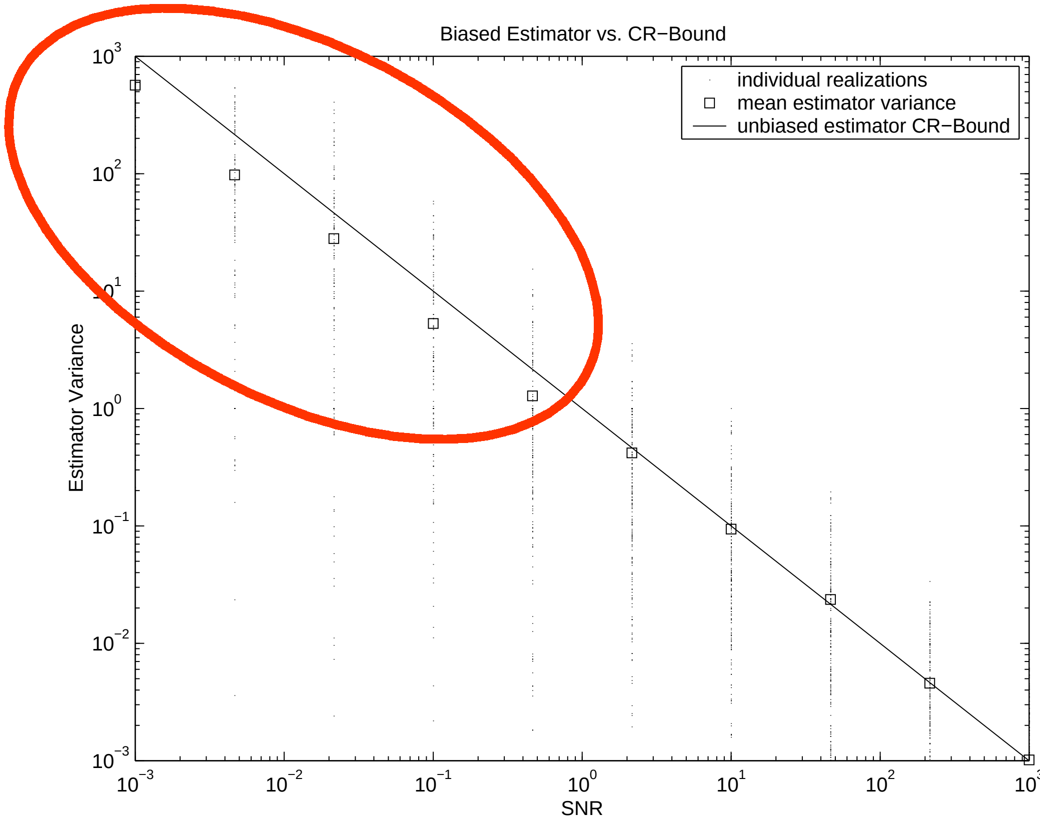
- Which results in ...

$$\hat{\theta}_{ML}(Y) = \begin{cases} Y & (Y > 0) \\ 0 & (Y \leq 0) \end{cases}$$

Numerical Simulation of new & improved Estimator

- Estimator variance is less then predicted by unbiased estimator CR-Bound
- Why?
- Estimator is *biased*

$$E\left[\hat{\theta}\right] - \theta \neq 0$$



Biased Estimator CR Bound

$$\mathrm{var}_{\underline{\theta}}\left(\hat{\theta}_p\right) \geq \left[\nabla m_{\underline{\theta}}\right]^T F_{\underline{\theta}}^{-1}\left[\nabla m_{\underline{\theta}}\right]$$

- Mean function $m_{\underline{\theta}} \equiv E_{\underline{\theta}}\left[\hat{\theta}_p\right]$
- Bias function $b_{\underline{\theta}} \equiv E_{\underline{\theta}}\left[\hat{\theta}_p\right] - \theta_p$
- Re-arrange & evaluate terms…

$$\mathrm{var}_{\underline{\theta}}\left(\hat{\theta}_p\right) \geq \left[\underline{e}_p + \nabla b_{\underline{\theta}}\right]^T F_{\underline{\theta}}^{-1}\left[\underline{e}_p + \nabla b_{\underline{\theta}}\right]$$

Returning to our example...

- For scalar case, biased-estimator CR Bound reduces to following form:

$$\mathrm{var}\left(\hat{\theta}\right) \geq \left(1 + \frac{\mathrm{db}(\theta)}{\mathrm{d}\theta}\right)^2 F_\theta^{-1}$$

- Estimator $\hat{\theta}_{ML}(Y) = Y \; (Y > 0)$

- Bias function $b_\theta\left(\hat{\theta}\right)$ = *complicated mess involving error functions, etc., etc.*

- Bias gradient $\dfrac{db_\theta}{d\theta}$ = *even more complicated mess involving error functions, etc., etc.*

So, what's the problem?

- Biased Estimator CR-Bound dependent on the estimator's bias-function
- Different estimator, different bias function
  - Bound is estimator specific
- How about estimator bias-function length or norm as a measure of overall bias ?
  - Remove estimator dependency

Uniform CR Bound for Point Source Estimation

- Given the Biased Estimator CR-Bound

$$\mathrm{var}_{\underline{\theta}}\left(\hat{\theta}_p\right) \geq \left[\underline{e}_p + \nabla b_{\underline{\theta}}\right]^T F_{\underline{\theta}}^{-1}\left[\underline{e}_p + \nabla b_{\underline{\theta}}\right]$$

- Find the lower bound among all possible estimator with a given bias gradient *length*

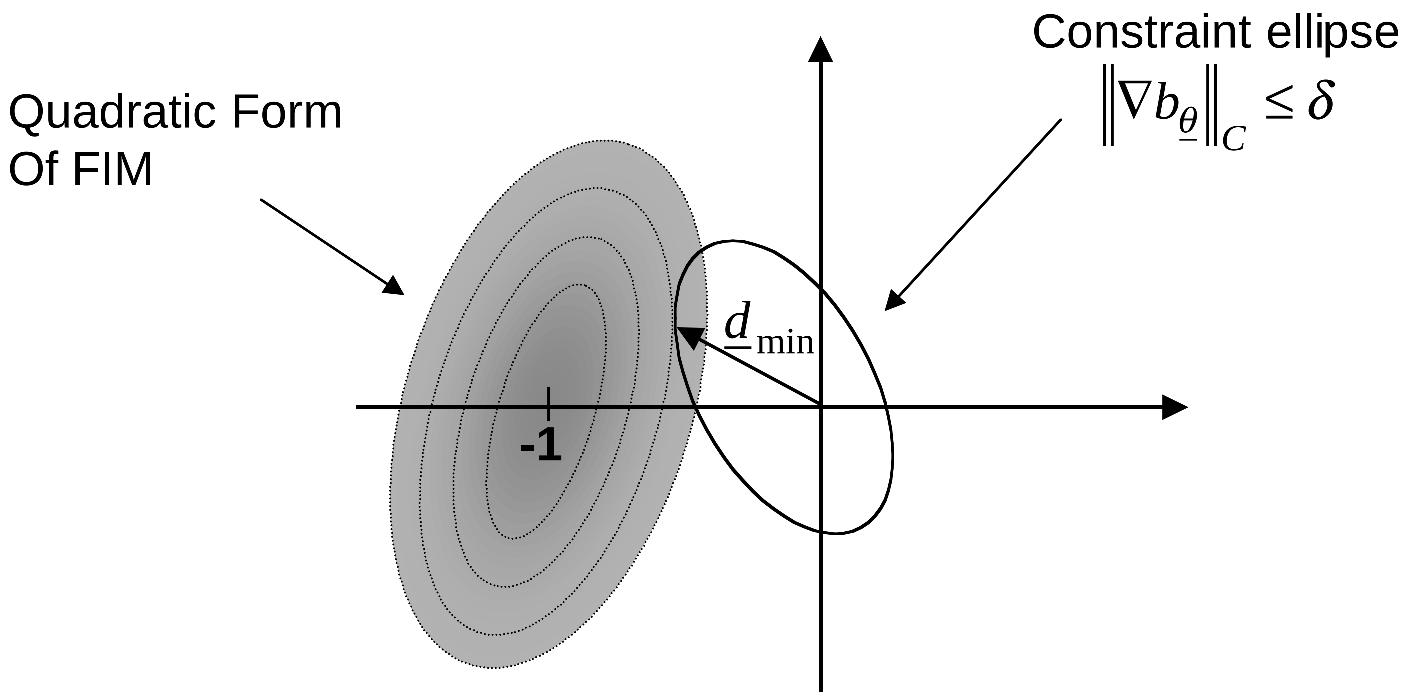
$$\left\|\nabla b_{\underline{\theta}}\right\|_C \leq \delta$$

- Pose problem as minimization of a quadratic form

$$\mathrm{var}_{\underline{\theta}}\left(\hat{\theta}_p\right) \geq \underset{\underline{d}\,:\,\|\underline{d}\|_C \leq \delta}{\text{arg - min}} \left[\underline{e}_p + \underline{d}\right]^T F_{\underline{\theta}}^{-1}\left[\underline{e}_p + \underline{d}\right]$$

Graphical Interpretation of UCRB for Point Source Estimation

$$\mathrm{var}_{\underline{\theta}}\left(\hat{\theta}_p\right) \geq \underset{\underline{d}\,:\,\|\underline{d}\|_C \leq \delta}{\text{arg - min}} \left[\underline{e}_p + \underline{d}\right]^T F_{\underline{\theta}}^{-1}\left[\underline{e}_p + \underline{d}\right]$$



What's the Deal with the Bias Gradient Norm?

- Maximum variation of the bias function over an ellipsoidal neighborhood about $\underline{\theta}$ is bounded by $\left\|\nabla b_{\underline{\theta}}\right\|_C$
- For point-source estimation, bias-function gradient is related to the point-response function
  - norm(bias-gradient) ~ norm(error in estimator psf)
  - Gives measure of overall error in estimator psf

  - *Hero/Fesser/Usman, "Exploring Estimator Bias-Variance Tradeoffs Using the Uniform CR Bound", IEEE-TSP, July '96*

Estimator Mean Response and Bias Functions

- Example: Linear Additive Gaussian Measurements
  - Measurement Equation $\underline{Y} \sim N\left(A\underline{\theta}, \Sigma\right)$
  - Fisher Information Matrix $F_Y = A^T \Sigma^{-1} A$
  - QPML Vector Estimator $\hat{\underline{\theta}}(\underline{Y}) = \left(F_Y + \beta P\right)^{-1} A \Sigma^{-1} \underline{Y}$
  - QPML Pixel Estimator $\hat{\theta}_p(\underline{Y}) = \underline{e}_p{}^T \hat{\underline{\theta}}(\underline{Y})$
  - Estimator Mean $m_{\underline{\theta}} = \underline{e}_p{}^T \left(F_Y + \beta P\right)^{-1} F_Y \underline{\theta}$
  - Estimator Bias $b_{\underline{\theta}} = \underline{e}_p{}^T \left[\left(F_Y + \beta P\right)^{-1} F_Y - I\right] \underline{\theta}$

Mean and Bias Gradient for Point Source Estimation

- Mean Estimator Response to Point Source

$$m_{\underline{\theta}} = \left(F_Y + \beta P\right)^{-1} F_Y \underline{e}_p$$

- Estimator Mean Gradient

$$\nabla m_{\underline{\theta}} = F_Y \left(F_Y + \beta P\right)^{-1} \underline{e}_p$$

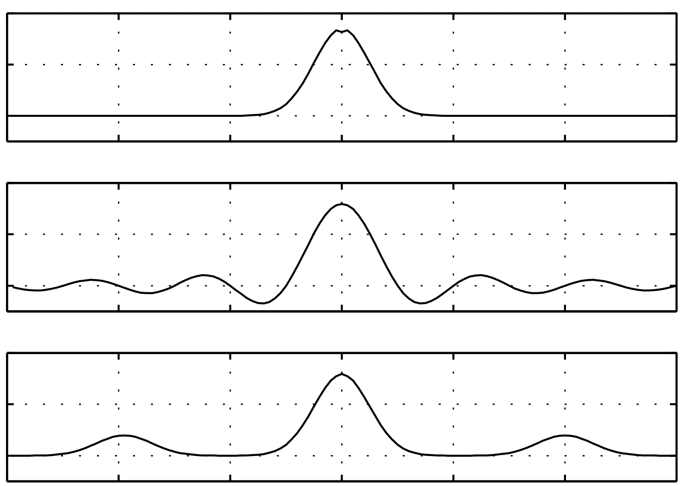
- Estimator Bias Gradient

$$\nabla b_{\underline{\theta}} = F_Y \left(F_Y + \beta P\right)^{-1} \underline{e}_p - \underline{e}_p$$

- *Gradients are measures of Point Response Function*

Problem with UCRB

- Multiple different estimator point response functions can have *identical* bias gradient length, but with widely different resolution properties



Extend the UCRB: Add a Resolution Measure

- Estimator mean gradient is related to the point response function
- Pose a resolution criteria on $\nabla m_{\underline{\theta}}$
  - *2nd-Moment relative to pth-pixel*

$$\gamma^2 = \frac{\sum_i (p-i)^2 \left(\nabla m_{\underline{\theta}}\right)_i^2}{\sum_i \left(\nabla m_{\underline{\theta}}\right)_i^2}$$

$$\gamma^2 = \frac{\nabla m_{\underline{\theta}}^T M_p \nabla m_{\underline{\theta}}}{\nabla m_{\underline{\theta}}^T \nabla m_{\underline{\theta}}} = \frac{\left\| \underline{e}_p + \nabla b_{\underline{\theta}} \right\|_{M_p}^2}{\left\| \underline{e}_p + \nabla b_{\underline{\theta}} \right\|_2^2}$$

Solving for Extended UCRB

- Perform same minimization as before

$$\text{var}_{\underline{\theta}}\left(\hat{\theta}_p\right) \geq \underset{\underline{d}}{\text{arg - min}} \left[\underline{e}_p + \underline{d}\right]^T F_{\underline{\theta}}^{-1}\left[\underline{e}_p + \underline{d}\right]$$

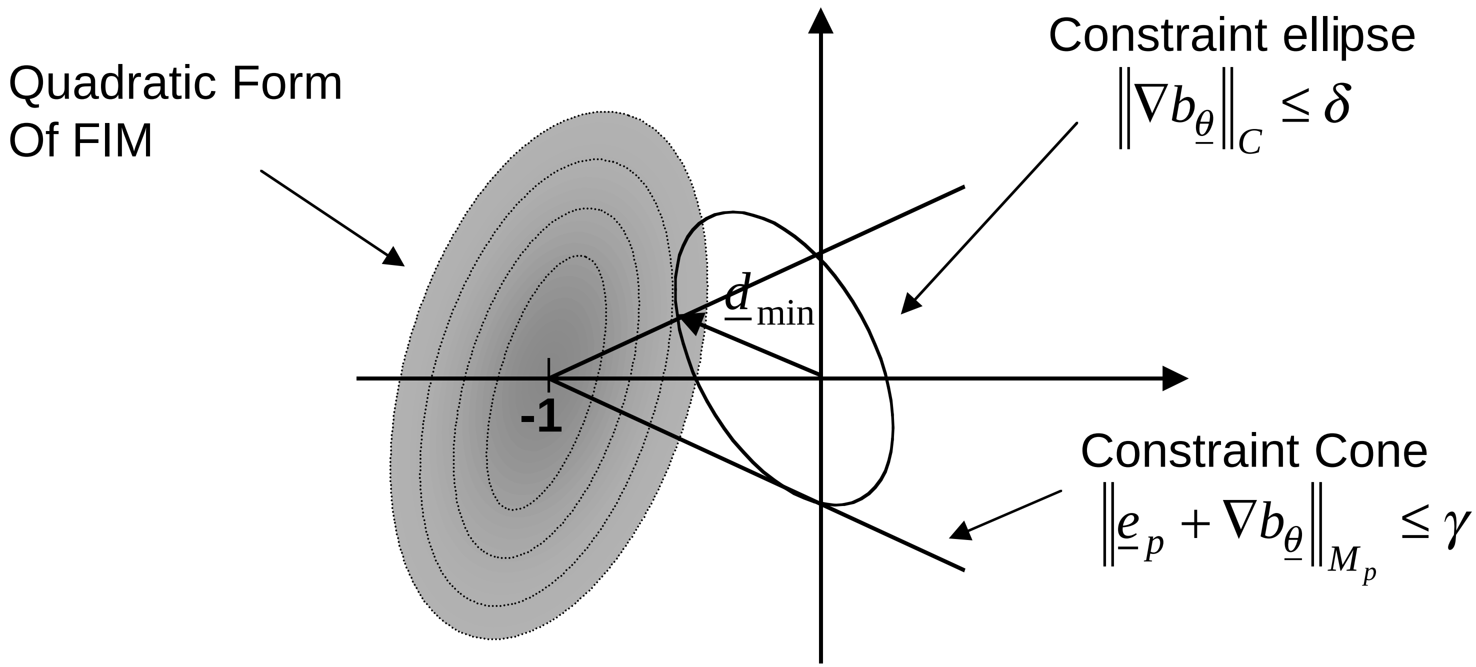
- Subject to the following two constraints:

  - 1) Bias Gradient Constraint $\left\|\nabla b_{\underline{\theta}}\right\|_C \leq \delta$
  - 2) Resolution Constraint $\left\|\underline{e}_p + \nabla b_{\underline{\theta}}\right\|_{M_p} \leq \gamma$

- Calculate resulting Bias-Resolution-Variance surface

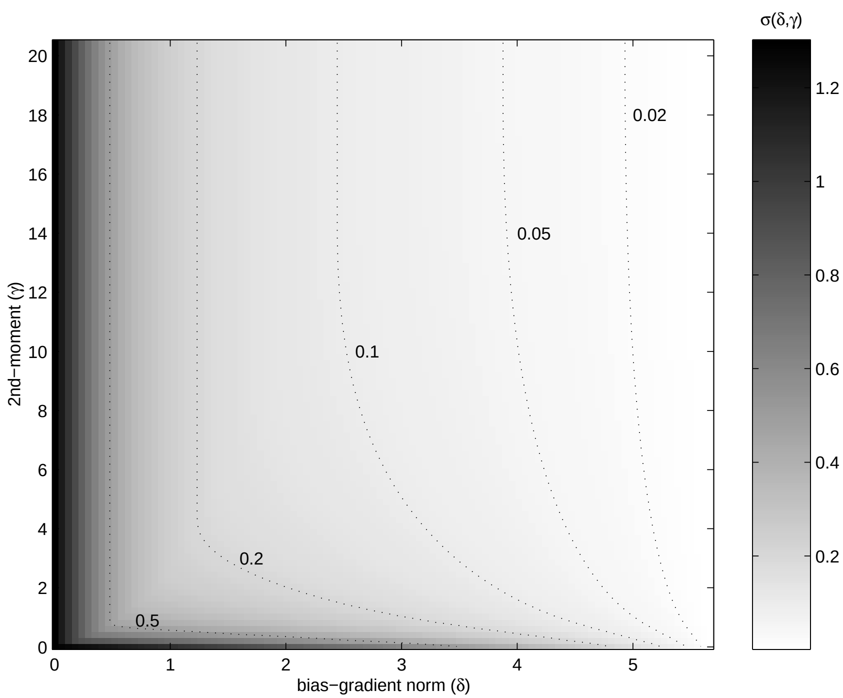
Graphical Interpretation of UCRB with Resolution Constraint

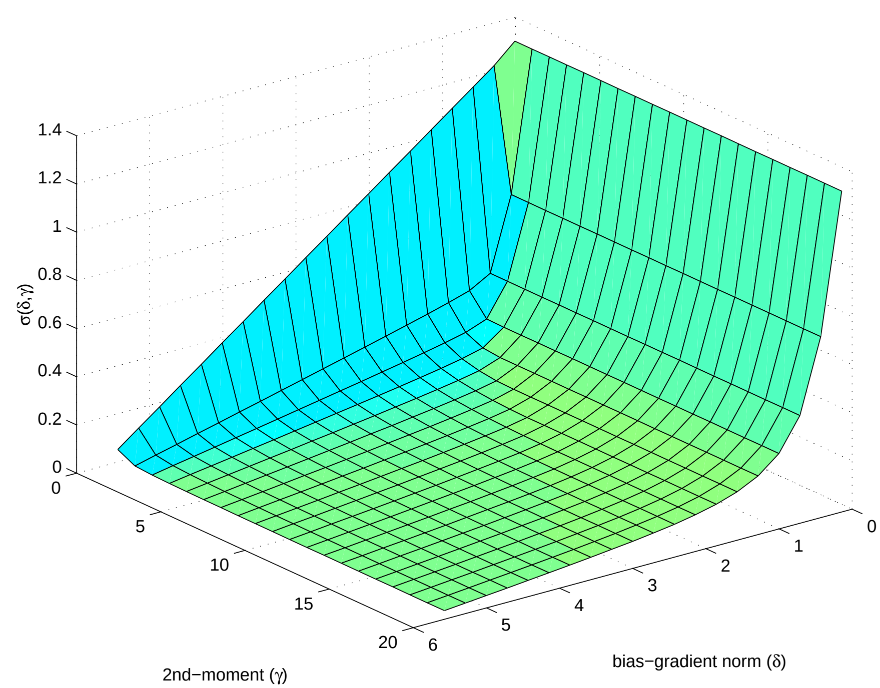
$$\mathrm{var}_{\underline{\theta}}\left(\hat{\theta}_p\right) \geq \underset{\underline{d}}{\text{arg - min}} \left[\underline{e}_p + \underline{d}\right]^T F_{\underline{\theta}}^{-1}\left[\underline{e}_p + \underline{d}\right]$$



Example UCRB Calculation with Resolution Constraint

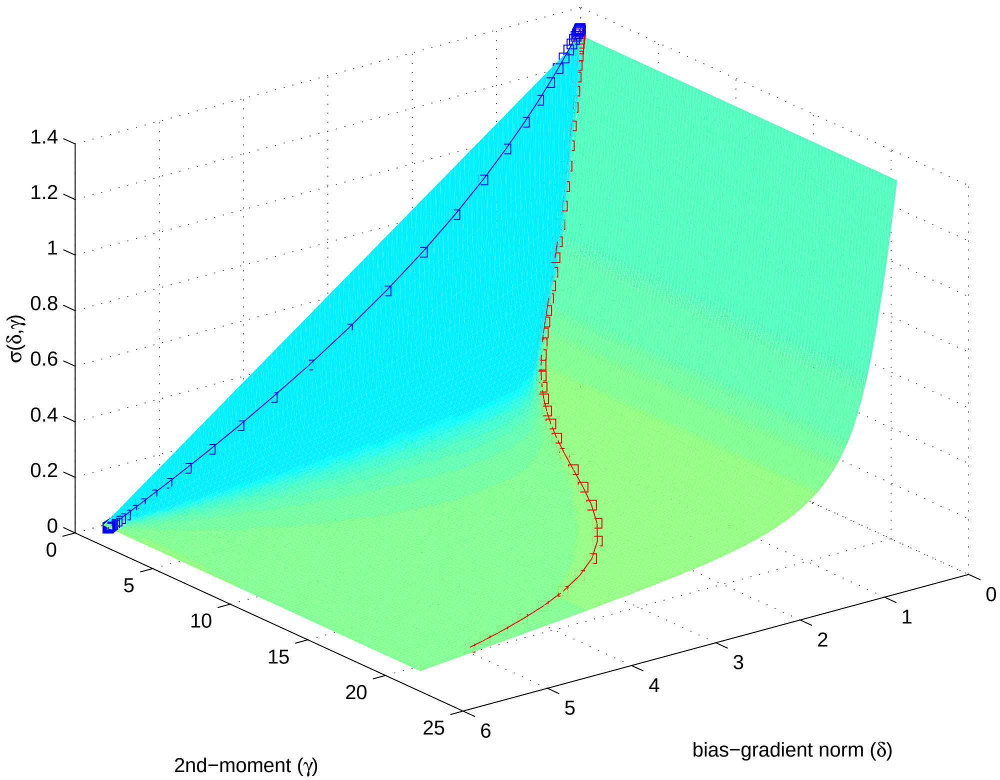
- Linear Additive Gaussian Inverse Problem
- Single Pixel Estimation Task





QPML-Estimator Variance vs. Extended UCRB

- Linear Additive Gaussian Inverse Problem
- Single Pixel Estimation Task
- Variance trajectories parameterized by smoothing penalty
  - Identity Penalty (blue)
  - Smoothing Penalty (red)



Example UCRB Calculation