Weiss-Weinstein Bounds for Various Priors

Florian Xaver*, Christoph F. Mecklenbräuker, Peter Gerstoft, and Gerald Matz

***Abstract*—We address analytic solutions of the Weiss-Weinstein bound (WWB), which lower bounds the mean squared error of Bayesian inferrers. The bound supports discrete, absolutely continuous, and singular continuous probability distributions, the latter corresponding to joint estimation and detection. We present new analytical solutions for truncated Gaussian, Laplace, categorical, uniform, and Bernoulli distributions. We focus on sparse signals modeled by a Laplace prior as used in Bayesian LASSO methods, priors of truncated Gaussian densities, and uninformative priors. In general, finding the tightest WWB of a model is a non-convex optimization problem. Hence, we show numerical examples of known and new WWBs to gain additional insight.**

***Index Terms*—Weiss-Weinstein lower bound, Bayesian inference, mean squared error, LASSO, categorical, uniform, Laplace, truncated Gaussian**

I. Introduction

In this paper we address discrete-time stochastic models in a Bayesian context. Random parameters are inferred from random observations. Estimation and detection (classification) are types of inference for continuous and discrete random parameters. They are designed by minimizing an expected loss function. If the loss function is the mean squared error, then the optimization result is the minimum mean squared error (MMSE) inferrer. Hence, the mean squared error of an inferrer is a performance indicator.

The study of lower bounds on the mean squared error of a Bayesian estimator [1]–[4] entails various bounds [5]–[8]. The most prominent bound is the Bayesian Cramér-Rao bound (CRB) for linear models, where the prior and the likelihood function are Gaussian. In this case, the mean squared error of the MMSE estimator coincides with the CRB.

We are interested in Bayesian lower bounds for the inference of parameters, which are jointly applicable to discrete and continuous random state variables. Additionally, the bound shall be valid for probability densities with finite support. It turns out that the regularity conditions for the applicability of the Bayesian CRB are too restrictive for discrete parameters [9], [10]. We want bounds with relaxed regularity conditions, which are applicable to discrete parameters, especially to discrete parameters that stems from quantized continuous states. This requirement guides us to the Weiss-Weinstein bounds (WWB) [5], [7], [11], [12] and their sequential versions [10], [13]–[15].

*Florian Xaver `florian@xaver.me`, Gerald Matz `gmatz@nt.tuwien.ac.at`, and Christoph Mecklenbräuker `cfm@nt.tuwien.ac.at` are with the Institute of Telecommunications, Vienna University of Technology, Gusshausstrasse 25/E389, 1040 Vienna, Austria. Peter Gerstoft `gerstoft@ucsd.edu` is with Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0238 , California.

The WWB is parametrized by a test point. The optimal test point provides the tightest WWB. For linear models and sample spaces $\mathbb{R}^N$, the WWB depends on the Bhattacharyya coefficient (BC) [10]. If the sample space is a subset of $\mathbb{R}^N$, then this leads to a generalized Bhattacharyya coefficient (GBC). Using analytic solutions of the (G)BC for popular distributions, numerical computations give insight into the (G)BC and WWB.

We open with a summary of WWBs in Section II, adding singular-continuous distributions that are inducing joint estimation and detection (e.g. in [15]). In Section III we present the BC stemming from Gaussian distributions and derive the GBC from categorical, Bernoulli, Laplace, uniform, and truncated Gaussian distributions. The newly derived WWBs are used as priors in Sections V to VII. A Laplacian prior induces sparsity and leads to the Bayesian least absolute shrinkage and selection operator (LASSO). The truncated Gaussian prior introduces a finite support as in practical problems, whereas a uniform prior is uninformative, and links to frequentist models.

II. Weiss-Weinstein bounds

Consider the probability space $(\mathbb{R}^N, \mathcal{B}, P_{\mathsf{x}})$ with the N-dimensional sample space $\mathbb{R}^N$, the induced Borel algebra $\mathcal{B}$, and the probability measure $P_{\mathsf{x}}(\mathcal{B}) = P(\mathsf{x} \in \mathcal{B})$ [16], [17].

Due to Lebesgue decomposition, a probability distribution may consists of an absolutely continuous, a singular (discrete), and a singular continuous part, i.e.

$$P_{\mathsf{x}} = c_1 P_{\mathsf{x}}^{\mathrm{c}} + c_2 P_{\mathsf{x}}^{\mathrm{d}} + c_3 P_{\mathsf{x}}^{\mathrm{s}} \ , \qquad c_1 + c_2 + c_3 = 1 \ .$$

Thus, the expectation of a measurable vector-valued function $\boldsymbol{g}(\boldsymbol{x})$ is

$$\begin{aligned} \mathrm{E}_{\mathsf{x}}\left(\boldsymbol{g}(\boldsymbol{x})\right) &:= \int_{\mathbb{R}^N} \boldsymbol{g}(\boldsymbol{x}) \mathrm{d}P_{\mathsf{x}}(\boldsymbol{x}) = c_1 \int_{\mathbb{R}^N} \boldsymbol{g}(\boldsymbol{x}) f_{\mathsf{x}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &\quad + c_2 \sum_{\boldsymbol{x} \in \mathcal{C} \subset \mathbb{R}^N} \boldsymbol{g}(\boldsymbol{x}) p_{\mathsf{x}}(\boldsymbol{x}) + c_3 \int_{\boldsymbol{x} \in \mathbb{R}^N} \boldsymbol{g}(\boldsymbol{x}) \mathrm{d}P_{\mathsf{x}}^{\mathrm{s}} \end{aligned}$$

with the probability density function (PDF) $f_{\mathsf{x}}(\boldsymbol{x}) = \mathrm{d}P_{\mathsf{x}}^{\mathrm{c}}(\boldsymbol{x}) / \mathrm{d}\,\lambda^{\mathrm{c}}(\boldsymbol{x})$ and the probability mass function (PMF) $p_{\mathsf{x}}(\boldsymbol{x}) = \mathrm{d}P_{\mathsf{x}}^{\mathrm{d}}(\boldsymbol{x}) / \mathrm{d}\,\lambda_{\mathcal{C}}^{\mathrm{d}}(\boldsymbol{x})$. If $P_{\mathsf{x}}^{\mathrm{s}} = P_{\mathsf{x}}^{\alpha_1} \times \ldots \times P_{\mathsf{x}}^{\alpha_N}$, $\alpha_i \in \{c, d\}$, is a product measure, then a *hybrid density* $v_{\mathsf{x}}(\boldsymbol{x})$ exists as a product of PDFs and PMFs. We call all three types simply probability density (PD) [18].

In what follows, we use the notation λ_{x} for probability measures whenever we assume the existence of a PD for the random variable $\boldsymbol{x}$. To simplify notation, we write $\mathrm{E}\left(\cdot\right) := \mathrm{E}_{\mathsf{x},\mathsf{y}}\left(\cdot\right)$, $f(\boldsymbol{x}) := f_{\mathsf{x}}(\boldsymbol{x})$, $p(\boldsymbol{x}) := p_{\mathsf{x}}(\boldsymbol{x})$, and $v(\boldsymbol{x}) := v_{\mathsf{x}}(\boldsymbol{x})$.

For the Bayesian estimate $\hat{\boldsymbol{x}}(\boldsymbol{y})$, the parameters $\boldsymbol{x} \in \mathbb{R}^N$ are inferred from measurements $\boldsymbol{y} \in \mathbb{R}^M$. In what follows, we consider the affine measurement equation

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{v} \ , \qquad \boldsymbol{x} \sim v(\boldsymbol{x}) \ , \qquad \boldsymbol{v} \sim v(\boldsymbol{v}) \ ,$$

with measurement matrix $\boldsymbol{C}:\mathbb{R}^N\mapsto\mathbb{R}^M$ and random noise $\boldsymbol{v}\in\mathbb{R}^N$. The resulting estimation error of estimate $\hat{\boldsymbol{x}}$ is defined by $\boldsymbol{e}:=\hat{\boldsymbol{x}}(\boldsymbol{y})-\boldsymbol{x}$. The mean squared error of the inferrer is lower bounded by the Weiss-Weinstein matrix $\boldsymbol{W}$ parametrized by a test-point matrix $\boldsymbol{H}=[\boldsymbol{h}_1,\ldots,\boldsymbol{h}_N]$. The test point $\boldsymbol{h}_a,\ a=1,\ldots,N$, pertains to element $[\boldsymbol{x}]_a$. For the bound, we insert $s_i=s_j={}^1\!/_2$ and (38) into (39)-(41) of [5] and reformulate it. Hence, the mean-squared-error matrix is bounded as (left-hand side minus right-hand side is positive semi-definite.)

$$\mathrm{E}\left(\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\right)\succcurlyeq\boldsymbol{W}(\boldsymbol{H}):=\boldsymbol{H}\boldsymbol{J}^{-1}\boldsymbol{H}^{\mathrm{T}}\ , \tag{1}$$

where $\boldsymbol{J}=\boldsymbol{J}(\boldsymbol{H})$ depends on $\boldsymbol{H}$ with typical element

$$[\boldsymbol{J}]_{a,b}:=\frac{\varrho(-\boldsymbol{h}_a,\boldsymbol{h}_b)-\varrho(\boldsymbol{h}_a,\boldsymbol{h}_b)}{\rho_{\mathsf{v}}(\boldsymbol{C}\boldsymbol{h}_a,\boldsymbol{0})\rho_{\mathsf{x}}(\boldsymbol{h}_a,\boldsymbol{0})\rho_{\mathsf{v}}(\boldsymbol{C}\boldsymbol{h}_b,\boldsymbol{0})\rho_{\mathsf{x}}(\boldsymbol{h}_b,\boldsymbol{0})}\ , \tag{2}$$

with the symmetrized version of the *generalized* non-metric Bayesian *Bhattacharyya coefficient*

$$\begin{aligned}\varrho(\boldsymbol{h}_a,\boldsymbol{h}_b):=[&\rho_{\mathsf{v}}(\boldsymbol{C}\boldsymbol{h}_a,\boldsymbol{C}\boldsymbol{h}_b)\rho_{\mathsf{x}}(\boldsymbol{h}_a,\boldsymbol{h}_b)\\&+\rho_{\mathsf{v}}(-\boldsymbol{C}\boldsymbol{h}_a,-\boldsymbol{C}\boldsymbol{h}_b)\rho_{\mathsf{x}}(-\boldsymbol{h}_a,-\boldsymbol{h}_b)]\ ,\end{aligned} \tag{3}$$

and the non-symmetric Bhattacharyya coefficient (GBC)

$$\rho_{\mathsf{v}}(\boldsymbol{h}_a,\boldsymbol{h}_b)=\int_{\mathcal{V}}\sqrt{v_{\mathsf{v}}(\boldsymbol{v}+\boldsymbol{h}_a)v_{\mathsf{v}}(\boldsymbol{v}-\boldsymbol{h}_b)}\mathrm{d}\lambda_{\mathsf{v}}\ , \tag{4}$$

where $\mathcal{V}=\{\boldsymbol{v}:v_{\mathsf{v}}(\boldsymbol{v})>0\}$. The GBC ρ_{x} is similar. Due to the product measure λ_{v}, the integral in (4) is an abbreviation of sums and Lebesgue integrals. The multiplications of the GBCs in (2) and (3) are due to linearity and statistical independence. Eq. (3) simplifies for symmetric PDs. For infinite support of $v_{\mathsf{v}}(\boldsymbol{v})$, (4) becomes the BC [19]–[21]

$$\rho_{\mathsf{v}}(\boldsymbol{h}_a,\boldsymbol{h}_b)=\int_{-\infty}^{\infty}\sqrt{v_{\mathsf{v}}(\boldsymbol{v}+\boldsymbol{h}_a)v_{\mathsf{v}}(\boldsymbol{v}-\boldsymbol{h}_b)}\mathrm{d}\lambda_{\mathsf{v}}\ . \tag{5}$$

Observe that if we substite $\boldsymbol{v}':=\boldsymbol{v}-\boldsymbol{h}_b$ in (4), then the GBC is still dependent on $\boldsymbol{h}_a$ and $\boldsymbol{h}_b$. We will see that the BC of the Gaussian and Laplace distributions depend only on the sum $\boldsymbol{h}:=\boldsymbol{h}_a+\boldsymbol{h}_b$ and that PDs with finite support induce box constrains on the test-point vector $\boldsymbol{h}$ [10].

III. Analytic Solutions

In what follows, we derive analytic solutions of the BC of several distributions. Observe from (4) that if $v_{\mathsf{v}}(\boldsymbol{v})=v_{\mathsf{v}_1}(v_1)\ldots v_{\mathsf{v}_N}(v_N)$, then $\rho_{\mathsf{v}}(\boldsymbol{h}_a,\boldsymbol{h}_b)=\rho_{\mathsf{v}_1}(h_{a,1},h_{b,1})\ldots\rho_{\mathsf{v}_N}(h_{a,N},h_{b,N})$. Thus, we consider univariate PDs whenever the elements of the parameter vector are assumed to be statistically independent.

A. Categorical Distribution

Categorical distributions are defined by a probability mass function. If the categorical distribution stems from the equidistant quantization (discretization) of a continuous distribution, $\ell\Delta_{\boldsymbol{v}}\in\mathbb{R}^N$, $\ell\in\mathcal{L}\subset\mathbb{Z}$, we can use (1) to (4). In this case, the mean squared error makes sense. Inserting its density

$$p(v)=\sum_{\ell\in\mathcal{L}}p_\ell\mathbb{1}_{v=\ell\Delta_v}$$

with probability masses p_ℓ and indicator function $\mathbb{1}_\cdot$, into (4) gives

$$\rho_{\mathsf{v}}^{\mathrm{C}}(h_a,h_b)=\sum_{v\in\mathcal{L}\Delta_v}\left[\sum_{\ell\in\mathcal{L}}p_\ell\mathbb{1}_{v+h_a=\ell\Delta_v}\sum_{\ell\in\mathcal{L}}p_\ell\mathbb{1}_{v-h_b=\ell\Delta_v}\right]^{1/2}.$$

Observe that the GBC depends on h_a and h_b. The Bernoulli distribution $p_0\mathbb{1}_{v=0}+p_1\mathbb{1}_{v=\Delta_v}$ and the discrete uniform distribution $\mathrm{Unif}\,\{r,s\}$ with $p_\ell=p$ and $|h_a\pm h_b|\le s-r+1$ are examples of categorical distributions.

Next, we address absolutely continuous distributions.

B. Uniform Distribution

Let us insert the continuous uniform distribution $\mathrm{Unif}\,\{r,s\}$,

$$f(v)=\frac{1}{s-r}\mathbb{1}_{r\le v\le s}\ ,$$

into (4), i.e.

$$\frac{1}{s-r}\int_r^s\mathbb{1}_{r\le v+h_a\le s}\mathbb{1}_{r\le v-h_b\le s}\mathrm{d}v\ .$$

We distinguish four cases:

$$\begin{cases}\int_r^{s-h_a}\mathbb{1}_{r+h_b\le v\le s+h_b}\mathrm{d}v\ , & h_a\ge0,h_b\ge0\ ,\\ \int_{r-h_a}^{s}\mathbb{1}_{r+h_b\le v\le s+h_b}\mathrm{d}v\ , & h_a\le0,h_b\le0\ ,\\ \int_r^{s-h_a}\mathbb{1}_{r+h_b\le v\le s+h_b}\mathrm{d}v\ , & h_a>0,h_b<0\ ,\\ \int_{r-h_a}^{s}\mathbb{1}_{r+h_b\le v\le s+h_b}\mathrm{d}v\ , & h_a<0,h_b>0\ .\end{cases}$$

Merging the first two and splitting the last two, the GBC becomes[1]

$$\rho_{\mathsf{v}}^{\mathrm{cU}}(h_a,h_b)=\begin{cases}1-\frac{|h|}{s-r}\ , & h_a\ge0,h_b\ge0 \text{ or } h_a\le0,h_b\le0\ ,\\ 1-\frac{h_a}{s-r}\ , & h_a>0,h_b<0,h_a>-h_b\ ,\\ 1+\frac{h_b}{s-r}\ , & h_a>0,h_b<0,h_a\le-h_b\ ,\\ 1+\frac{h_a}{s-r}\ , & h_a<0,h_b>0,h_a>-h_b\ ,\\ 1-\frac{h_b}{s-r}\ , & h_a<0,h_b>0,h_a\le-h_b\ ,\\ 0\ , & \text{else}\ .\end{cases}$$

with $|h_a\pm h_b|<s-r$.

C. Laplace Distribution

The Laplace distribution $\mathrm{La}\,\{b,\mu_x\}$ is a sparse prior. Inserting Laplace's density $f(v)=\frac{1}{2b}e^{-\frac{|v-\mu_v|}{b}}$ into (5) and substituting $w=v-\mu_v$ and $h=h_a+h_b$ gives

$$\frac{1}{2b}\int_{-\infty}^{\infty}e^{-\frac{1}{2b}(|w+h|+|w|)}\mathrm{d}w\ .$$

If $h>0$, then the integral splits into three terms,

$$\begin{aligned}w>0\ ,w+h>0\ &:\quad \frac{1}{2b}\int_0^\infty e^{-w/b-h/2b}\mathrm{d}w=\frac{1}{2}e^{-h/2b}\ ,\\ w<0\ ,w+h<0\ &:\quad \frac{1}{2b}\int_{-\infty}^{-h}e^{w/b-h/2b}\mathrm{d}w=\frac{1}{2}e^{-h/2b}\ ,\\ w<0\ ,w+h>0\ &:\quad \frac{1}{2b}\int_{-h}^{0}e^{-h/2b}\mathrm{d}w=\frac{h}{2b}e^{-h/2b}\ .\end{aligned}$$

[1] This is a revised version of BC in [10], which contains only the first case.

Due to symmetry, for $h \leq 0$, there are three similar cases. The composition of the terms provides the BC

$$\rho_{\mathsf{v}}^{\mathrm{L}}(h_a, h_b) = \rho_v^{\mathrm{L}}(h) = \begin{cases} \left(1 + \frac{h}{2b}\right) e^{-h/2b} \ , & h > 0 \ , \\ \left(1 - \frac{h}{2b}\right) e^{h/2b} \ , & h \leq 0 \ . \end{cases}$$

D. (Truncated) Gaussian Distribution

The density of Gaussian distribution $\mathrm{N}\left\{\boldsymbol{\mu_v}, \boldsymbol{C_v}\right\}$ is

$$v(\boldsymbol{v}) := \frac{1}{(2\pi)^{N/2} \left(\det \boldsymbol{C}_{\mathsf{v}}\right)^{1/2}} e^{-\frac{1}{2}\|\boldsymbol{v}-\boldsymbol{\mu_v}\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}} \ .$$

with the mean $\boldsymbol{\mu_v}$, the covariance matrix $\boldsymbol{C}_{\mathsf{v}}$, and the weighted norm $\|\boldsymbol{h}\|_{\boldsymbol{C}_{\mathsf{v}}^{-1}} := (\boldsymbol{h}^{\mathrm{T}}\boldsymbol{C}_{\mathsf{v}}^{-1}\boldsymbol{h})^{1/2}$. Then the Bayesian BC [10] is given by

$$\rho_{\mathsf{v}}^{\mathrm{G}}(\boldsymbol{h}_b, \boldsymbol{h}_b) = \rho_{\mathsf{v}}^{\mathrm{G}}(\boldsymbol{h}) := e^{-\frac{1}{8}\|\boldsymbol{h}\|^2_{\boldsymbol{C}_v^{-1}}} \ .$$

The truncated Gaussian $\mathrm{N}\left\{\mu_{\mathsf{v}}, \boldsymbol{C_v}, \boldsymbol{r}, \boldsymbol{s}\right\}$ is a Gaussian distribution limited to the support $[\boldsymbol{r}, \boldsymbol{s}]$, i.e., the density is given by

$$f(\boldsymbol{v}) = \frac{1}{c} e^{-\frac{1}{2}\|\boldsymbol{v}-\boldsymbol{\mu_v}\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}} \mathbb{1}_{\boldsymbol{r}\leq\boldsymbol{v}\leq\boldsymbol{s}} \ .$$

with appropiate normalization factor c. Inserting $f(\boldsymbol{v})$ into (4), substituting $\boldsymbol{v}' = \boldsymbol{v} - \boldsymbol{\mu}_{\mathsf{v}}$, and use normalization factor c, we get

$$\begin{aligned} &\frac{1}{c}\int_{\boldsymbol{r}-\boldsymbol{\mu}_{\mathsf{v}}}^{\boldsymbol{s}-\boldsymbol{\mu}_{\mathsf{v}}} e^{-\frac{1}{4}\left(2\|\boldsymbol{v}'\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}} + 2\boldsymbol{v}'^{\mathrm{T}}\boldsymbol{C}_{\mathsf{v}}^{-1}\boldsymbol{h}_a + \|\boldsymbol{h}_a\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}\right)} \\ &\times e^{-\frac{1}{4}\left(-2\boldsymbol{v}'^{\mathrm{T}}\boldsymbol{C}_{\mathsf{v}}^{-1}\boldsymbol{h}_b + \|\boldsymbol{h}_b\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}\right)} \\ &\times \mathbb{1}_{\boldsymbol{r}-\boldsymbol{h}_a\leq\boldsymbol{v}\leq\boldsymbol{s}-\boldsymbol{h}_a} \mathbb{1}_{\boldsymbol{r}+\boldsymbol{h}_b\leq\boldsymbol{v}\leq\boldsymbol{s}+\boldsymbol{h}_b} \mathrm{d}\boldsymbol{v}' \ . \end{aligned}$$

Next, we substitute $\boldsymbol{v}'' = \boldsymbol{v}' + {}^{\boldsymbol{h}_a}\!/_2 - {}^{\boldsymbol{h}_b}\!/_2$ and obtain

$$\begin{aligned} \rho_{\mathsf{v}}^{\mathrm{tG}}(\boldsymbol{h}_a, \boldsymbol{h}_b) =& \frac{1}{c}\rho_{\mathsf{v}}^{\mathrm{G}}(\boldsymbol{h}_a, \boldsymbol{h}_b) \times \int_{\boldsymbol{r}-\boldsymbol{\mu}_{\mathsf{v}}+\boldsymbol{h}_a/2-\boldsymbol{h}_b/2}^{\boldsymbol{s}-\boldsymbol{\mu}_{\mathsf{v}}+\boldsymbol{h}_a/2-\boldsymbol{h}_b/2} e^{-\frac{1}{2}\|\boldsymbol{v}''\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}} \\ &\times \mathbb{1}_{\boldsymbol{r}\leq\boldsymbol{v}''+\boldsymbol{h}_a/2+\boldsymbol{h}_b/2\leq\boldsymbol{s}} \mathbb{1}_{\boldsymbol{r}\leq\boldsymbol{v}''-\boldsymbol{h}_a/2-\boldsymbol{h}_b/2\leq\boldsymbol{s}} \mathrm{d}\boldsymbol{v}'' \ . \end{aligned}$$

The two cases $\boldsymbol{h}_a \geq 0, \boldsymbol{h}_b \geq 0$ and $\boldsymbol{h}_a \leq 0, \boldsymbol{h}_b \leq 0$ lead to

$$\frac{\rho_{\mathsf{v}}^{\mathrm{G}}(\boldsymbol{h}_a, \boldsymbol{h}_b)}{c} \int_{\boldsymbol{r}-\boldsymbol{\mu}_{\mathsf{v}}+|\boldsymbol{h}_a|/2+|\boldsymbol{h}_b|/2}^{\boldsymbol{s}-\boldsymbol{\mu}_{\mathsf{v}}-|\boldsymbol{h}_a|/2-|\boldsymbol{h}_b|/2} e^{-\frac{1}{2}\|\boldsymbol{v}''\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}} \mathrm{d}\boldsymbol{v}'' \ .$$

If $\boldsymbol{h}_a \leq 0, \boldsymbol{h}_b > 0$, then $\rho_{\mathsf{v}}^{\mathrm{tG}}(\boldsymbol{h}_a, \boldsymbol{h}_b)$ is

$$\frac{1}{c}\rho_{\mathsf{v}}^{\mathrm{G}} \int_{\boldsymbol{r}-\boldsymbol{\mu}_{\mathsf{v}}+|\boldsymbol{h}_a|/2-|\boldsymbol{h}_b|/2}^{\boldsymbol{s}-\boldsymbol{\mu}_{\mathsf{v}}-|\boldsymbol{h}_a|/2-|\boldsymbol{h}_b|/2} e^{-\frac{1}{2}\|\boldsymbol{v}''\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}} \mathrm{d}\boldsymbol{v}''$$

for $\boldsymbol{h}_a + \boldsymbol{h}_b \leq 0$ and

$$\frac{1}{c}\rho_{\mathsf{v}}^{\mathrm{G}} \int_{\boldsymbol{r}-\boldsymbol{\mu}_{\mathsf{v}}+|\boldsymbol{h}_a|/2+|\boldsymbol{h}_b|/2}^{\boldsymbol{s}-\boldsymbol{\mu}_{\mathsf{v}}+|\boldsymbol{h}_a|/2-|\boldsymbol{h}_b|/2} e^{-\frac{1}{2}\|\boldsymbol{v}''\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}} \mathrm{d}\boldsymbol{v}''$$

for $\boldsymbol{h}_a + \boldsymbol{h}_b > 0$. If $\boldsymbol{h}_a > 0, \boldsymbol{h}_b \leq 0$, then $\rho_{\mathsf{v}}^{\mathrm{tG}}(\boldsymbol{h}_a, \boldsymbol{h}_b)$ is

$$\frac{1}{c}\rho_{\mathsf{v}}^{\mathrm{G}} \int_{\boldsymbol{r}-\boldsymbol{\mu}_{\mathsf{v}}+|\boldsymbol{h}_a|/2-|\boldsymbol{h}_b|/2}^{\boldsymbol{s}-\boldsymbol{\mu}_{\mathsf{v}}+|\boldsymbol{h}_a|/2+|\boldsymbol{h}_b|/2} e^{-\frac{1}{2}\|\boldsymbol{v}''\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}} \mathrm{d}\boldsymbol{v}''$$

for $\boldsymbol{h}_a + \boldsymbol{h}_b \leq 0$ and

$$\frac{1}{c}\rho_{\mathsf{v}}^{\mathrm{G}} \int_{\boldsymbol{r}-\boldsymbol{\mu}_{\mathsf{v}}+|\boldsymbol{h}_a|/2-|\boldsymbol{h}_b|/2}^{\boldsymbol{s}-\boldsymbol{\mu}_{\mathsf{v}}-|\boldsymbol{h}_a|/2-|\boldsymbol{h}_b|/2} e^{-\frac{1}{2}\|\boldsymbol{v}''\|^2_{\boldsymbol{C}_{\mathsf{v}}^{-1}}} \mathrm{d}\boldsymbol{v}''$$

for $\boldsymbol{h}_a + \boldsymbol{h}_b > 0$. Due to the finite support, $|\boldsymbol{h}_a \pm \boldsymbol{h}_b| \leq \boldsymbol{s} - \boldsymbol{r}$. This integral and Factor c in general can only be numerically evaluated.

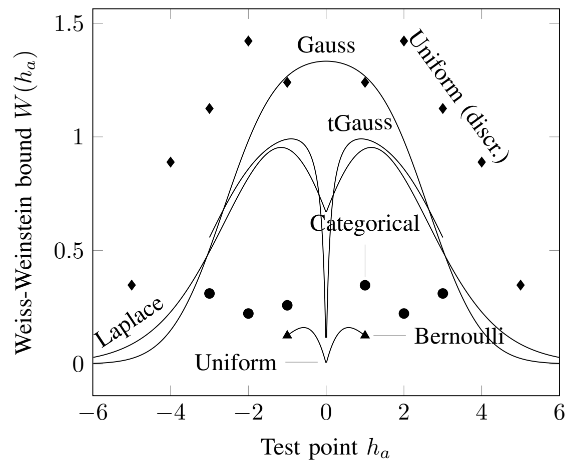

Figure 1. Weiss-Weinstein bound vs. test point for equal prior and noise distributions. $h_a = h_b$ for univariate distributions.

IV. Numerical Examples

In the sequel, we use the scalar measurement model with $C = 1$. Test matrix $\boldsymbol{H}$ shrinks to a scalar, i.e. $h_a = h_b$ in (2). Fig. 1 plots WWBs for Laplace $\mathrm{La}\left\{1, 0\right\}$, Gaussian $\mathrm{N}\left\{0, 2\right\}$, uniform $\mathrm{Unif}\left\{-1, 1\right\}$, and truncated Gaussian $\mathrm{N}\left\{0, 2, -3, 3\right\}$ prior and measurement distributions.

In general, the WWB is non-concave regarding the test point and no general solution for the optimal test point h exists. The existence of an analytic solution of the tightest bound depends on the BC of the particular distributions. However, due to symmetry, for unimodal densities only $h \geq 0$ is needed and the BC is quasi-concave. The optimal test point is small.

In case of Gaussian distributions, the optimal test point tends to $h_a \to 0$. The WWB approaches the CRB, cf. [10] for a proof. The tightest WWB in the Gaussian case coincides with the CRB and is reached by an optimal estimator.

Fig. 1 also shows the WWB of the categorical distribution $\left\{0.5, 0.2, 0.1, 0.2\right\}$, the Bernoulli distribution $\left\{0.5, 0.5\right\}$, and the discrete uniform distribution $\mathrm{Unif}\left\{-3, 3\right\}$ for $\Delta_v = 1$.

The shape of WWB of discrete distributions follow that of continuous distributions. In case of Bernoulli distributions, there are only two possible test points $h_a = h_b = \pm\Delta_v = \pm 1$. According to [10], the WWBs of continuous distributions and their quantized discrete versions are equal for the same test points. Comparing the WWB of the discrete uniform distribution with the bound of the continuous uniform distribution, we observe how a smaller support lowers the Weiss-Weinstein bound.

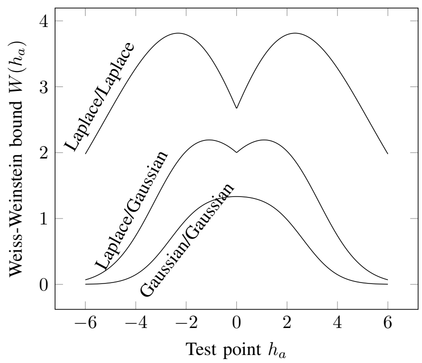

Figure 2. WWB for a Laplace prior and Gaussian likelihood compared to pure Laplacian and pure Gaussian distributions.

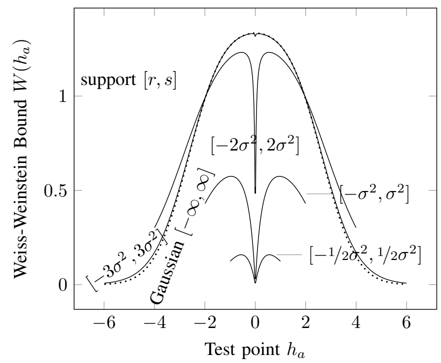

Figure 3. WWB for truncated Gaussian densities of support $[r, s]$ vs. test point. The Gaussian curve is for comparison (dotted).

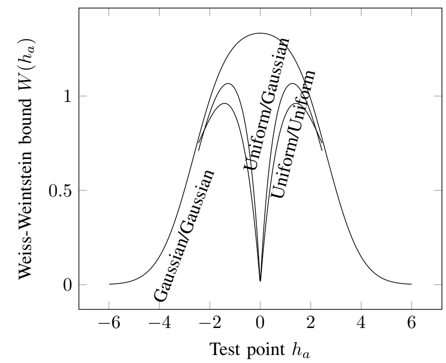

Figure 4. WWB for a uniform prior and Gaussian likelihood function compared to pure uniform and pure Gaussian case.

V. Sparse Signals – Laplace Prior

The least absolute shrinkage and selection operator (LASSO) is a prominent estimation method for sparse signals $\boldsymbol{x}$ [22]. LASSO minimizes the squared error $\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{C}\boldsymbol{x}\|_2^2$, where vectors $\boldsymbol{x}, \boldsymbol{y}$ are deterministic. Assuming a sparse signal $\boldsymbol{x}$, it makes sense to constrain the ℓ_1-norm $\|\boldsymbol{x}\|_1 = \sum_n |x_n| < b \in \mathbb{R}_+$. The associated Lagrangian becomes

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{C}\boldsymbol{x}\|_2^2 + b' \|\boldsymbol{x}\|_1 \ , \qquad b' \in \mathbb{R}_+.$$

In the Bayesian context $\boldsymbol{x}$ and $\boldsymbol{y}$ are random. Then the minimization is interpreted as the maximization of the posterior distribution, if the likelihood function is Gaussian and the prior is Laplacian.

In Fig. 2, we show the WWB for the scalar linear measurement model for pure Laplace distributions $x, v \sim \mathrm{La}\{0, 2\}$, pure Gaussian distributions $x, v \sim \mathrm{N}\{0, 2\}$ and Laplace prior / Gaussian noise $x \sim \mathrm{La}\{0, 2\}, v \sim \mathrm{N}\{0, 2\}$. Observe that the WWB of the mixed case lies in-between the pure examples. In the limit $h \to 0$, the bound is greater than zero.

VI. The Truncated Gaussian Density

Real-world applications that are modeled by a Gaussian prior, are usually box constrained. This leads to the truncated Gaussian distribution. Let $x, v \sim \mathrm{N}\{0, 2, r, s\}$. Fig. 3 shows the WW for different $[r, s]$. The Gaussian case ($r \to -\infty$ and $s \to \infty$) is our reference.

Compare the curves with that of continuous uniform distributions in Fig. 1. They look similarly. The bound decreases with decreasing support size $s - r$. Compare it with the pure uniform curve in Fig. 3. The support of the truncated Gaussian density defines, whether the WWB tends to the uniform or to the Gaussian WWB.

VII. Uninformative Prior – Uniform Prior

A uniform prior is uninformative in the sense that it only defines a minimum and maximum value. Measurement y updates the beliefe of signal x. Due to the uninformative prior, the noise distribution has a strong impact on the believe update.

Let us now compare uniform prior / Gaussian noise $x \sim \mathrm{Unif}\{-2.45, 2.45\}, v \sim \mathrm{N}\{0, 2\}$ with uniform distributions $x, v \sim \mathrm{Unif}\{-2.45, 2.45\}$, and Gaussian distributions $x, v \sim \mathrm{N}\{0, 2\}$. For the uniform distribution we used

$$-r = s = +1/2\sqrt{12\sigma^2} \approx 2.45$$

to achieve identical variance $\sigma^2 = 2$ for the Gaussian and uniform distribution.

Next, we plot the WWB for the scalar linear measurement model in Fig. 4. Replacing the uniform by a Gaussian likelihood function increases the lower bound. Observe that the WWB of the mixed model lies above the pure uniform case, but has a similar shape. Two characteristics influence the maximum WWB: the support and the flatness of the densities. The greater the support, the greater the possible error. The flatter the density, the greater the error.

VIII. Conclusion

The Weiss-Weinstein bound (WWB) is parametrized by test points leading to different levels. The WWB depends on the density's variability and on its support. They apply to both discrete and absolutely continuous probability distributions. This is useful for quantized variables, where the values in the sample space have a physical interpretation. Furthermore, the WWB supports singular-continuous multivariate probability distributions with a hybrid joint density. A hybrid joint density is the product of probability densities and probability mass functions. The WWB is applicable to probability distributions that are non-differentiable or have finite support, e.g. uniform or truncated Gaussian distributions, and distributions defined only on real positive support, e.g. exponential distributions. For linear models, the WWB depends on the generalized Bhattacharyya coefficient. If the distribution is unimodal, then the WWB is bimodal.

To exemplify its use, we discussed a linear measurement model. This illustrates the application of the WWB for sparse signal models with a Laplace-distributed prior and additive Gaussian noise. The WWB for a truncated Gaussian prior tends to the normally or uniformly distributed case depending on its support. The WWB of a model with a continuous uniform prior and a Gaussian likelihood function is greater than of the uniform-only model and has a similar shape.

References

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. NJ: Pearson Education, 1993, vol. 1.

[2] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle filters for tracking applications*. Boston: Artech House, 2004.

[3] O. Hlinka, O. Slučiak, F. Hlawatsch, P. M. Djurić and, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4334–4349, Aug. 2012.

[4] O. Hlinka, F. Hlawatsch, and P. M. Djurić, "Distributed particle filtering in agent networks: A survey, classification, and comparison," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 61–81, 2013.

[5] A. J. Weiss and E. Weinstein, "A general class of lower bounds in parameter estimation," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 338–342, 1988.

[6] H. L. Van Trees and K. L. Bell, *Bayesian bounds for parameter estimation and nonlinear filtering/tracking*. IEEE Press, 2007.

[7] A. Renaux, P. Forster, P. Larzabal, C. D. Richmond, and A. Nehorai, "A fresh look at the Bayesian bounds of the Weiss-Weinstein family," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5334–5352, Nov. 2008.

[8] D. A. Koshaev and O. A. Stepanov, "Application of the Rao-Cramer inequality in problems of nonlinear estimation," *Journal of Comp. and System Sci. Int.*, vol. 36, pp. 220–227, Apr. 1997.

[9] Z. Duan, V. P. Jilkov, and X. R. Li, "State estimation with quantized measurements: Approximate MMSE approach," in *11th Int. Conf. on Inform. Fusion*, IEEE, 2008, pp. 1–6.

[10] F. Xaver, P. Gerstoft, G. Matz, and C. Mecklenbräuker, "Analytic sequential Weiss-Weinstein bounds," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 5049–5062, 2013.

[11] N. D. Tran, A. Renaux, R. Boyer, S. Marcos, and P. Larzabal, "Weiss-Weinstein bound for MIMO radar with colocated linear arrays for SNR threshold prediction," *Signal Processing*, vol. 92, no. 5, pp. 1353–1358, 2012.

[12] A. Renaux, "Weiss–Weinstein bound for data-aided carrier estimation," *Signal Process. Letters, IEEE*, vol. 14, no. 4, pp. 283–286, 2007.

[13] I. Rapoport and Y. Oshman, "Recursive Weiss-Weinstein lower bounds for discrete-time nonlinear filtering," in *43rd IEEE Conf. on Decision and Control*, vol. 3, Dec. 2004, pp. 2662–2667.

[14] S. Reece and D. Nicholson, "Tighter alternatives to the Cramer-Rao lower bound for discrete-time filtering," in *8th Int. Conf. on Inform. Fusion*, vol. 1, Jul. 2005, 6 pp.

[15] F. Xaver, G. Matz, P. Gerstoft, and C. F. Mecklenbräuker, "Predictive state vector encoding for decentralized field estimation in sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Kyoto, JP, Mar. 2012, pp. 2661–2664.

[16] P. Billingsley, *Probability and measure*, 3rd. John Wiley & Sons, Inc., New York, 2012.

[17] F. E. Burk, *A Garden of Integrals*, 1st ed., ser. The Dolciani Math. Expositions 31. The Math. Assoc. of Am., 2007.

[18] F. Xaver, "Mixed discrete-continuous Bayesian inference: Censored measurements of sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5609–5620, Nov. 2015.

[19] Z. Ben-Haim and Y. C. Eldar, "A comment on the Weiss–Weinstein bound for constrained parameter sets," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4682–4684, 2008.

[20] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc*, vol. 35, no. 99-109, p. 4, 1943.

[21] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. 15, no. 1, pp. 52–60, Feb. 1967.

[22] T. Park and G. Casella, "The Bayesian Lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008. eprint: http://dx.doi.org/10.1198/016214508000000337.