Abstract—It is shown that the generalized Gaussian distribution maximizes the generalized Cramer-Rao (CR) bound for the $p$th absolute central moment of any classical location parameter unbiased estimator. The underlying maximization is taken over the class of distributions with fixed and finite $p$th-order moment and exhibits particular utility in minimax designs as well as in worst-case performance analysis. The relationship between the generalized Gaussian density and the generalized CR bound is further examined for the model of a mixture of generalized Gaussian distributions as well as for scenarios where multiple independent generalized Gaussian observations are involved.

I. INTRODUCTION

Lower bounds on the performance of classical parameter estimation procedures have been the subject of research work for the past several decades. The bounds derived so far were mainly lower bounds on the variance of such procedures [1]–[4], [6].

In this correspondence, we examine lower bounds on the $p$th absolute central moment of unbiased parameter estimators. Restricting ourselves to the unbiased class, we immediately obtain bounds on the $p$th-order estimation error, which is a useful performance measure that penalizes increasingly large deviations between the transmitted signal and the estimate with increasing values of $p$.

There are two major reasons that justify the importance of lower bounds on the $p$th absolute central moments of estimators. First, for given estimators, it may be difficult to compute their moments directly, or at least, it may be a computationally intensive task. It may rather be easier to derive lower bounds on these moments for any estimator within a large class (e.g., class of unbiased schemes) and then see how the moments of the given estimator compare with these lower bounds. Second, there might be cases where we want to design a scheme that minimizes the $E[(\hat{\theta} - \theta)^p]$ for $p > 2$, where $\hat{\theta}$ is an unbiased estimator of a parameter $\theta$. Since the loss function $[\hat{\theta} - \theta]^p$, as $p$ increases, pays more attention to the extreme values of the estimator and, hence, to outlying observations, the performance of the resulting scheme is strongly influenced by the tails of the data distribution. It is thus implied that such a criterion might be adequate if we want to penalize heavily the large deviations between the transmitted signal and the estimate. However, the requirement $p > 2$ usually leads to mathematically intractable solutions or prohibitively complex formulas for the resulting estimators, and adaptive solutions are often unavoidable [7]. All in all, in situations where the adopted criterion for either the design or the performance evaluation of an estimation scheme is the $p$th-order mean error, it is useful to search for tight lower bounds on the pertinent criterion to which the performance of any estimator could be compared.

A general class of Bayesian (random parameter) lower bounds on the moments of the estimation error is derived in [11]. In this correspondence, we consider a generalization of the Cramer-Rao (CR) bound on the $p$th-order absolute error of classical location parameter unbiased estimators, i.e., estimators where their location is modeled as a deterministic unknown parameter. This bound, which comes as a natural generalization of the regular CR bound, is derived here in a very simple way, although it can be derived more elaborately as a special case of the Barankin bound [1]. The main contribution of this work is the relationship between the generalized CR bound and the generalized Gaussian density. This result, which is presented in Theorem 2, is a generalization of the relationship between the original (second-order) CR bound and the Gaussian density that has not been mentioned in the literature, at least to the best of the authors’ knowledge. This work also examines the case where we have a mixture of generalized Gaussian densities. Situations with multiple independent generalized Gaussian observations are studied as well. For the latter cases, a looser bound is also provided since explicit evaluation of the generalized CR bound might be a difficult task.

II. GENERALIZED CR BOUND AND THE LOCATION PARAMETER CASE

Theorem 1: Let $p(x|\theta) > 0$ for every $x$ in the sample space $X$ and $\theta \in \Theta$, where $\Theta$ is an interval on the real line with distinct endpoints, i.e., $\Theta$ has one of the following forms: $(a, b)$, $(a, b]$, $[a, b)$, or $[a, b]$ for some $a, b$ such that $a < b$. The sample space $X$ is assumed to be independent of the parameter $\theta$. In addition, let $(d/d\theta)p(x|\theta) = \int f(d/d\theta)\tilde{g}(x|\theta)p(x|\theta)\,dx$, and let $f(p(x|\theta))$ be differentiable under the integral sign. Then, for any unbiased estimator of a real parameter $\theta$ and $p \geq 1$

$$E_\theta [|\hat{\theta} - \theta|^p] \geq \int \left[ \left| \frac{d}{d\theta} p(x|\theta) \right| p(x|\theta) \right]^{1/p} \cdot \sgn \left[ \frac{d}{d\theta} \log p(x|\theta) \right] \,dx. \tag{1}$$

1/p + 1/q = 1, with equality iff there are nonnegative constants $\mu$ and $\nu$ depending on $\theta$, not both zero, such that

$$\mu (\hat{\theta} - \theta) = \nu \left( \frac{\hat{\theta} - \theta}{\log p(x|\theta)} \right)^{1/p} \cdot \sgn \left( \frac{d}{d\theta} \log p(x|\theta) \right).$$

where

$$E_\theta \left[ \left| \frac{d}{d\theta} \log p(x|\theta) \right|^{1/p} \cdot \sgn \left( \frac{d}{d\theta} \log p(x|\theta) \right) \right] = 0.$$

In addition, the general form of $p(x|\theta)$ that satisfies (1) with equality is the following:

$$p(x|\theta) = q(x) \exp \left[ - \int_0^\theta (\mu/\nu)^{1/q} \frac{d}{du} (|\hat{\theta} - u|^p) \,du \right].$$

Proof:

$$\int_X p(x|\theta)\,dx = 1 \Rightarrow \int_X \theta \frac{d}{d\theta} \log p(x|\theta)p(x|\theta)\,dx = 0$$

and

$$\int (\hat{\theta} - \theta) \frac{d}{d\theta} \log p(x|\theta)p(x|\theta)\,dx = 1.$$

Subtracting the above equations, taking absolute values, and applying H"{o}lder’s inequality (the same inequality was also used in [11]), we obtain

$$\int (\hat{\theta} - \theta)^p p(x|\theta)\,dx \geq \left\{ \left( \int \frac{d}{d\theta} \log p(x|\theta) \right)^p \right\}^{1/p} \cdot \left( \int p(x|\theta)\,dx \right)^{-1/q}. $$
Hölder’s inequality is satisfied with equality [9] iff there are nonnegative constants \( \alpha, \beta \) not both zero such that
\[
\alpha |\hat{g}(x) - \theta|^\gamma = \beta \left| \frac{d}{d\theta} \log p(x|\theta) \right|^\gamma, \quad \text{and}
\]
\[
\text{sgn} (\hat{g}(x) - \theta) = \text{sgn} \left[ \frac{d}{d\theta} \log p(x|\theta) \right]
\]
which implies
\[
\mu \{ \hat{g}(x) - \theta \} = \nu \left| \frac{d}{d\theta} \log p(x|\theta) \right|^\gamma \text{sgn} \left[ \frac{d}{d\theta} \log p(x|\theta) \right]
\]
where
\[
\mu = \alpha^{1/\gamma}, \quad \nu = \beta^{1/\gamma}
\]
or equivalently
\[
(\mu/\nu)^{1/\gamma} |\hat{g}(x) - \theta|^{\gamma/\gamma} \text{sgn} [\hat{g}(x) - u] = \frac{d}{du} \log p(x|u).
\]
Thus
\[
p(x|\theta) = q(x) \exp \int_\theta^x (\mu/\nu)^{\gamma/\gamma} |\hat{g}(x) - \theta|^{\gamma/\gamma} \text{sgn} (\hat{g}(x) - \theta) d\theta
\]
where \( q(x) = \exp \{ h(x) \} \) is a function of \( x \) chosen to give \( f(p(x|\theta) dx = 1 \). However
\[
\frac{d}{d\theta} |\hat{g}(x) - \theta|^{\gamma} = -p |\hat{g}(x) - \theta|^{\gamma-1} \text{sgn} (\hat{g}(x) - \theta)
\]
and since \( 1/p + 1/q = 1 \), which implies that \( q = p/(p-1) \) and \( p-1 = p/q \), we obtain
\[
p(x|\theta) = q(x) \exp \left[ -\int_\theta^x p^{-1}(\mu/\nu)^{\gamma/\gamma} \frac{d}{du} \left| \hat{g}(x) - u \right|^{\gamma} du \right]. \]

In the following, we are going to show that the denominator of the right-hand side of (1) is convex with respect to \( p(x) \). This is essentially a corollary to the following lemma, and it is applicable to the general parameter estimation problem. Next, using Corollary 1, we find the density that maximizes the bound given by (1) for the location parameter case. Theorem 2 presents the latter result.

**Lemma 1:** Let \( v_1 > 0, v_2 > 0, u_1 > 0, u_2 > 0, 0 < a < 1 \). Then
\[
\frac{[av_1 + (1-a)u_1]^a}{[av_1 + (1-a)v_2]^a} \leq a \left( \frac{u_1}{v_1} \right)^a + (1-a) \left( \frac{u_2}{v_2} \right)^a.
\]

**Proof:** Set \( v = av_1 + (1-a)v_2, \beta = av_1/v \). Then \( 0 < \beta < 1 \) and since \( q > 1 \)
\[
\frac{[av_1 + (1-a)u_2]^a}{[av_1 + (1-a)v_2]^a} = \left( \frac{uv_1}{v_1} + (1-\beta) \left( \frac{u_2}{v_2} \right)^a \right)^a 
\]
\[
\leq \beta \left( \frac{u_1}{v_1} \right)^a + (1-\beta) \left( \frac{u_2}{v_2} \right)^a 
\]
\[
= \alpha \left( \frac{u_1}{v_1} \right)^a + (1-a) \left( \frac{u_2}{v_2} \right)^a.
\]
where in the above proof, we used the convexity of the function \( x^a, x \geq 0, q > 1 \). \( \square \)

**Corollary 1:** If we apply Lemma 1 for
\[
\alpha_1 = |p_1'(x|\theta)|, \quad \alpha_2 = |p_2'(x|\theta)|,
\]
\[
v_1 = p_1(x|\theta), \quad v_2 = p_2(x|\theta), \quad q > 1
\]
we obtain
\[
\int \frac{|p_1'(x|\theta) + (1-a)p_2'(x|\theta)|^\gamma}{|p_1(x|\theta) + (1-a)p_2(x|\theta)|^\gamma} dx 
\]
\[
\leq \int \frac{|p_1'(x|\theta)|^\gamma}{|p_1(x|\theta)|^\gamma} dx + (1-a) \int \frac{|p_2'(x|\theta)|^\gamma}{|p_2(x|\theta)|^\gamma} dx
\]
where
\[
p_1'(x|\theta) = \frac{d}{d\theta} p_1(x|\theta) \quad \text{and} \quad p_2'(x|\theta) = \frac{d}{d\theta} p_2(x|\theta). \quad \square
\]

**Theorem 2:** Let \( P \) be the class of densities \( f \) that are a.e. differentiable. They have fixed and finite absolute moment of order \( p \), \( p \geq 1 \) (i.e., \( \int_{-\infty}^{\infty} |x|^p f(x) dx = M_p < \infty \) \( \forall f \in P \) and are positive \(( f(x) > 0 \) for every \( x \) in the sample space. Then, the density that maximizes the right-hand side of (1) for the location parameter case is given by
\[
f_p(x) = \frac{p^{(\nu-1)/\nu}}{2\alpha(\nu-1)} \exp \left\{ \frac{1}{p} \frac{|x|^\nu}{p(\nu-1)} \right\}
\]
which is a generalized Gaussian density of order \( p \) with zero mean and \( \nu \)th absolute moment \( M_p \) equal to \( \nu^{\nu-1}/\nu \).

**Proof:** If \( \theta \) is a location parameter, then there exists a density \( f(y) \) such that \( p(x|\theta) = f(x-\theta) \), and for \( y = x-\theta \), the generalized CR bound can be written as
\[
E_p [ |\hat{g}(x|\theta) - \theta|^\nu ] 
\]
\[
\geq \int \left[ \frac{d}{dy} f(y) \right] f(y)^{p-1} f(y) dy 
\]
\[
= 1 \left[ \int \left[ \frac{d}{dy} f(y) \right] f(y)^{(p-1)} dy \right]^{(p-1)}.
\]
Thus, as for the case of \( p = q = 2 \) in the class of location parameter distributions, the bound given by (2) is independent of the parameter.
In addition, the class \( P \) is convex.

Therefore
\[
\int_{-\infty}^{\infty} y \frac{d}{dy} f(y) dy = y f(y)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(y) dy
\]
\[
= -1 \int y \left( \frac{d}{dy} \log f(y) \right) f(y) dy 
\]
\[
= 1.
\]
By Hölder’s inequality
\[
\left\{ \int |y|^p f(y) dy \right\}^{1/p} \left\{ \int \left| \frac{d}{dy} \log f(y) \right|^q f(y) dy \right\}^{1/q} \geq 1
\]
or equivalently
\[
\left\{ \int \left| \frac{d}{dy} \log f(y) \right|^q f(y) dy \right\}^{p/q} \geq \left\{ \int |y|^p f(y) dy \right\}^{-1} = (M_p)^{-1}
\]
with equality iff there exists a positive constant \( \alpha \) such that
\[
|\alpha y|^p = \left| \frac{d}{dy} \log f(y) \right|^q \quad \text{and} \quad \text{sgn}(y) = -\text{sgn} \left[ \frac{d}{dy} f(y) \right].
\]
The above equality conditions show that
\[
\frac{d}{dy} \log f(y) = \frac{d}{dy} \left[ -\alpha^{1/q} (p/q + 1)^{-1} |y|^{(p/q)+1} + H \right] \Rightarrow f(y)
\]
\[
= K \exp \left\{ -\alpha^{1/q} ((p/q + 1)^{-1} |y|^{(p/q)+1} \right\}
\]
where \( K = e^H \) is a normalizing constant. Since \( 1/p + 1/q = 1 \),
we have that \( p/q = p - 1 \), \( (p/q + 1) = p + 1 \), and \( 1/q = (p - 1)/p \).
Therefore, \( f(y) = K \exp \left\{ -p^{-2} \alpha^{1/(p-1)} |y|^{p/p} \right\} \). The constant \( K \)
can be calculated if we use the requirement that \( f(y) dy = 1 \). Thus
\[
K = \left( \int_{-\infty}^{\infty} \exp \left\{ -p^{-1} \alpha^{(p-1)/p} |y|^{p/p} \right\} dy \right)^{-1}.
\]
If we use the formula
\[
\int_{0}^{\infty} x^{\nu-1} e^{-\mu x} dx = |\mu|^{-\nu/p} \Gamma(\nu/p), \mu, \nu > 0
\]
then for \( \nu = 1 \) and \( \mu = p^{-1} \alpha^{(p-1)/p} \), we obtain
\[
K = \left\{ 2p^{-1} \alpha^{(p-1)/p} \right\}^{-1/2} \Gamma(1/p) p^{-1}
\]
\[
= p^{-1} \alpha^{(p-1)/p} \Gamma(1/p)^{-1}
\]
where
\[
\Gamma(n) = \int_{0}^{\infty} x^{n-1} e^{-x} dx.
\]
Therefore, the density for which the bound in (2) becomes maximum
is the generalized Gaussian \( f_p(x) \) of order \( p \) with mean 0 and \( p \)th
absolute moment \( \alpha^{(p-1)/p} \), i.e.,
\[
f_p(x) = \frac{p^{(p-1)/p}}{2 \alpha^{(p-1)/p} \Gamma(1/p)} \exp \left\{ -\frac{\alpha^{(p-1)/p}}{p} |x|^p \right\}
\]
and the maximum value the bound attains equals the \( p \)th absolute
moment \( M_p \) of the generalized Gaussian density, which is \( \alpha^{(p-1)/p} \).
We note that for \( p = 2 \), we have \( \alpha^{(1-p)/p} = \sigma^2, \Gamma(1/2) = \sqrt{\pi} \),
and \( f_2(x) \) is the Gaussian density \( N(0, \sigma^2) \), whereas for \( p = 1 \)
and \( \alpha^{(1-p)/p} = \mu \), we have the Laplace distribution with variance
parameter \( \mu \).

The bound given by (1) can be considered as the reciprocal of a
\( p \)th-order generalization of the Fisher information measure. As such,
Theorem 2 gives the worst performance of the best estimator (if it
exists) when the ”bestness” criterion is the \( p \)th-order estimation error.
Theorem 2 shows that the worst performance is attained by the
\( p \)th-order generalized Gaussian density. That is, the \( p \)th-order
generalized Gaussian density is the least favorable density within the class
of distributions with fixed and finite \( p \)th-order absolute central moment,
and thus, it attains minimal information. This result is particularly
useful when there is uncertainty in the statistical description of the
data expressed by a nonparametric class of distributions and any other
effort to assess the performance of an estimation scheme in terms of
the \( p \)th-order error fails. In addition, this maximization property of
the generalized Gaussian distribution can be of particular interest in
minimax designs. For example, one may wish to determine the \( M \)
estimator that minimizes the supremum of the \( p \)th-order absolute error
where the supremum is taken with respect to the density of the data.
Theorem 2 gives the least favorable density in the convex class \( P \).
If an estimator \( \hat{\theta} \) is minimax when the density is restricted to \( P \)
and \( \hat{\theta} \in P \) with equal suprema of the \( p \)th-order error taken over all
elements in \( P \), and \( P \), respectively, then \( \hat{\theta} \) is also minimax when the
density is permitted to vary over \( P \).

The generalized Gaussian density of order \( p \) with absolute second
moment \( \sigma^2 \) is given by
\[
f_p(x) = \frac{p}{2 \alpha(p) \Gamma(1/p)} \exp \left\{ -\frac{\alpha(x/A(p))^p}{p} \right\}
\]
\[\text{Fig. 1. Generalized Gaussian density of order } p \text{ with } 0\text{-mean and absolute second moment (variance) } \sigma^2 = 1.\]

\[\text{Fig. 2. Second-order CR bound for generalized Gaussian densities of order } p \text{ with variance } \sigma^2 = 1. \text{ The maximum value is attained for } p = 2.\]

[5], [8], where
\[
A(p) = \left[ \sigma^2 \Gamma(1/p) / \Gamma(3/p) \right]^{1/2}
\]
This is a class of symmetric unimodal density functions parametrized
by the variance \( \sigma^2 \) and the rate of exponential decay \( p > 0 \). Thus,
the above class consists of densities that include the normal Gaussian as
a special case (\( p = 2 \)) as well as those with relatively much faster
(\( p > 2 \)) or much slower (\( p < 2 \)) rates of exponential decay of their
tails.

\[\text{Example 1: From Theorem 2, among all densities with prespecified second moment } \sigma^2, \text{ the density that maximizes the second-order CR bound given by (2) is the generalized Gaussian of order } 2, \text{ and this maximum value equals } \sigma^2. \text{ Indeed, a simple calculation shows that the expectation } E_{f_2} \left[ |x|^p \right] \text{ for } p = 2 \text{ equals } \]
\[
\sigma^2 \Gamma(1/2) / \Gamma(3/2)^{1/2} = \sigma^2.
\]
\[\text{Fig. 1 shows the generalized Gaussian density with zero mean and absolute second-moment (variance) } \sigma^2 = 1 \text{ of different orders } p = 1, 1.5, 2, 3, 4. \text{ Fig. 2 shows the value of the second-order CR bound for different generalized Gaussian densities (with variance } \sigma^2 = 1) \text{ as a function of their order } p > 1. \text{ Indeed, the maximum value is attained for } p = 2. \]

In the following example, we consider a mixture of generalized
Gaussian densities that results in the actual density \( f \). In such a case,
the evaluation of the generalized CR bound can be a formidable task.
Therefore, a looser but easier to evaluate lower bound can be proven to be very useful. In this context, it is important to notice that if \[ p(x|\theta) = \sum_{i=1}^{n} \lambda_i p_i(x|\theta), \quad \sum_{i=1}^{n} \lambda_i = 1, \]
and \( \theta \) is a location parameter, then the following inequality holds for the generalized CR bound of order \( p \):
\[
\left\{ \frac{d}{dx} \log p(x|\theta) \right\}^p p(x|\theta) dx \geq \left\{ \sum_{i=1}^{n} \lambda_i \left( \int \frac{d}{dx} \log f_i(x) f_i(x) dx \right)^p \right\}^{-1}
\] (3)
where \( f_i(x - \theta) = p_i(x|\theta). \) This can be proved by a convexity argument on the denominator of the left-hand side of the previous inequality and by taking into account that \( \theta \) is a location parameter.

Example 2: Let \( f(x) = \lambda f_1(x) + (1 - \lambda) f_2(x), \quad 0 < \lambda < 1. \)

Case 1: Let \( f_1(x) = \frac{c}{2A(c) \Gamma(1/c)} e^{\left\{ -\left[ \frac{|x|^c}{A(c)^{1/c}} \right] \right\}}, \quad c > 2 \)
and \( f_2(x) = \frac{2}{2A(2) \Gamma(1/2)} e^{\left\{ -\left[ \frac{|x|^{2/c}}{A(2)^{1/c}} \right] \right\}}, \quad c = 2 \)
where \( A(c) = [\sigma^2 \Gamma(1/c)/\Gamma(3/c)]^{1/c}. \)

For \( c = 2 \), we have \( A(c) = A(2), \) i.e., \( f_2 \sim N(0, \sigma^2) \). Suppose also that we want to find the generalized CR bound of order \( p = 2 \) (this implies that \( q = 2 \)). From Theorem 2, we have that
\[
\int f_2(x) dx = \sigma^{-2}.
\]

On the other hand
\[
\int x f_2(x) dx = \left\{ \frac{1}{2} \Gamma(2/c - 1)/(\sigma^2 \Gamma(3/c)) \right\}^{-1} \sigma^{-2}.
\]
Therefore, from (3)
\[
\left\{ \frac{d}{dx} \log f(x) \right\}^2 f(x) dx = \left\{ \frac{1}{\sigma^2 \Gamma(1/c)/\Gamma(3/c)} \left( \frac{\Gamma((2c - 1)/c)}{\Gamma(1/c)} + \frac{1 - \lambda}{\sigma^2} \right) \right\}^{-1}. (4)
\]

In Fig. 3, we plot the second-order CR bound (the regular CR bound) of (4) for the Case 1 of Example 2 with \( \sigma^2 = 1 \). We see that the bound is maximized when we have no contamination at all (\( c = 2 \)). For \( c \neq 2 \), that is, when a Gaussian distribution is contaminated with a \( c \)-order generalized Gaussian, then the level \( \lambda \) of contamination increases (i.e., the influence of the regular Gaussian decreases), the bound decreases.

Case 2: Let \( f_1(x) \) and \( f_2(x) \) be as in Case 1. Suppose that we want to evaluate the \( c \)-order CR bound, which is given by (3) for \( p = c \). This implies that \( q = c/(c - 1) \) and \( c/q = c - 1 \). From Theorem 2
\[
\left\{ \frac{d}{dx} \log f(x) \right\}^p f(x) dx = \left\{ \int_{-\infty}^{\infty} |x|^{c/(c - 1)} f_1(x) dx \right\}^{-1} = \Gamma(1/c)/\left\{ \sigma^2 \Gamma(1/c)/\Gamma(3/c) \right\}^{1/c} \Gamma((c + 1)/c).
\]

On the other hand
\[
\left\{ \frac{d}{dx} \log f_2(x) \right\}^p f_2(x) dx = \left\{ \int_{-\infty}^{\infty} |x|^{c/(2c - 1)} f_2(x) dx \right\}^{-1} = 2^{c/2} \left\{ \frac{\Gamma((2c - 1)/(2c - 2))^{-1} \Gamma((2c - 1)/c)}{\sqrt{\pi}} \right\}^{-1} \sigma^{-2}.
\]

In Fig. 4, we plot the \( c \)-order CR bound for the case where a Gaussian density is mixed with a \( c \)-order generalized Gaussian, both with variance \( \sigma^2 = 1 \). For \( c = 2 \) (no contamination), the second-order CR bound is controlled by the variance \( \sigma^2 = 1 \) of the Gaussian distribution. For \( c \neq 2 \), the \( c \)-order CR bound is controlled by the \( c \)-order generalized Gaussian distribution. Thus, the bound increases as the influence of the \( c \)-order generalized Gaussian distribution increases (i.e., when \( \lambda \) increases) and reaches its maximum when only the \( c \)-order generalized Gaussian distribution is present (\( \lambda = 1 \)).

The next theorem deals with multiple independent observations.

Theorem 3: Let \( x = [x_1, x_2, \ldots, x_n]^T \) be a vector of \( n \) independent random variables, each having a density function \( p(x_i|\theta) \) parametrized by the same single parameter \( \theta \). In addition, let the assumptions of Theorem 1 hold for each of the \( p(x_i|\theta) \) and for the joint density \( p(x|\theta) \). Then
\[
E\left\{ |\hat{\theta}(x) - \theta|^p \right\} \geq \left\{ \int p(x|\theta) \left[ \prod_{i=1}^{n} p(x_i|\theta) d\theta \right]^{p/q} \right\}^{-1/c}.
\]
we have Fig. 4. cth-order gen. CR bound with $\sigma_1^2 = \sigma_2^2 = 1$. For $c = 2$ (no contamination) the second-order gen. CR bound is controlled by the cth-order gen. Gaussian density (it is maximum when only the cth-order gen. Gaussian density is present, i.e., when $\lambda = 1$).

and a looser bound is given by

$$E_0 \{ \hat{y}(x) - \theta^p \} \geq \left\{ \sum_{i=1}^{n} \left[ \int \frac{d}{dx} \log p(x_i | \theta) \right]^p p(x_i | \theta) dx_i \right\}^{-1/p}.$$ 

If $\{x_i\}$ are also identically distributed with density $p(x_i | \theta)$ and $\theta$ is a location parameter, then

$$E_0 \{ \hat{y}(x) - \theta^p \} \geq n^{-p} \left\{ \int \frac{d}{dx} \log p(x) \right\}^{-p} \int p(x) dx_1 \right\}^{-1/p}$$

which shows, as with expression (2), that for a location parameter $\theta$, the bound given by the above expression does not depend on $\theta$.

Proof: Applying Theorem 1 for the vector

$$x = [x_1, x_2, \ldots, x_n]^T$$

we have

$$\left\{ \int_{\mathbb{R}^n} \left[ \int \frac{d}{dx} \log p(x | \theta) \right]^p p(x | \theta) dx \right\}^{-1/p} \geq \left\{ \int_{\mathbb{R}^n} \sum_{i=1}^{n} \frac{d}{dx} \log p(x_i | \theta) \right\}^{-1/p} \left\{ \int \prod_{i=1}^{n} p(x_i | \theta) dx_i \right\}^{1/p}. \quad (5)$$

Then, Minkowski’s inequality gives

$$\left\{ \int_{\mathbb{R}^n} \sum_{i=1}^{n} \frac{d}{dx} \log p(x_i | \theta) \right\}^{1/p} \left\{ \int \prod_{i=1}^{n} p(x_i | \theta) dx_i \right\}^{1/p} \leq \sum_{i=1}^{n} \left\{ \int \frac{d}{dx} \log p(x_i | \theta) \right\}^{1/p} \left\{ \int p(x_i | \theta) dx_i \right\}^{1/p}.$$ 

Thus, a looser bound to (5) is given by

$$E_0 \{ \hat{y}(x) - \theta^p \} \geq \left\{ \sum_{i=1}^{n} \left[ \int \frac{d}{dx} \log p(x_i | \theta) \right]^p p(x_i | \theta) dx_i \right\}^{-1/p} \quad (6)$$

If $x_i$ are also identically distributed with density $p(x_i | \theta)$, then for a location parameter $\theta$, the bound becomes

$$E_0 \{ \hat{y}(x) - \theta^p \} \geq n^{-p} \left\{ \int \frac{d}{dx_1} \log p(x_1 | \theta) \right\}^{-p} \left\{ \int p(x_1 | \theta) dx_1 \right\}^{-1/p}.$$

Fig. 5 shows the value of the pth-order CR bound given by the above expression as a function of the sample size $n$ for $N(0, 1)$ i.i.d. data.

We observe that for $p = q = 2$, the regular CR inequality gives a tighter bound ($n$ in place of $n^q$). This is so because the Fisher information measure is additive for $p = 2$, whereas this is not true for $p \neq 2$.

Example 3: Let $n$, be i.i.d variables with density

$$f(x) = \frac{2}{2A(2)\Gamma(1/2)} \exp \left\{ -\left[ \frac{|x|}{A(2)} \right]^2 \right\}$$

where

$$A(2) = [\sigma^2 \Gamma(1/2) \Gamma(3/2)]^{1/2}$$

i.e., their distribution is $N(0, \sigma^2)$. Let us also assume that we observe $y_i = \theta + n_i$. Then, for any unbiased estimator $\hat{y}$ of $\theta$ that is based on the observations $y_1, \ldots, y_n$, it is true that

$$E_0 \{ \hat{y}(y) - \theta^p \} \geq n^{-p} \left\{ \int \frac{d}{dx} \log f(x) \right\}^{-p} \left\{ \int f(x) dx \right\}.$$ 

From Example 2 (Case 2), it follows that

$$E_0 \{ \hat{y}(y) - \theta^p \} \geq \frac{1}{n^p} \left( \frac{A(2)}{A(2)} \right)^{-p} \left( \frac{\Gamma(3/2)}{\Gamma(1/2)} \right)^{-p/2} \left( \frac{\Gamma(1/2)}{\Gamma(1/2)} \right)^{-p/2}.$$ 

Fig. 5 shows the value of the pth-order CR bound given by the above expression as a function of the sample size $n$ and when the data are Gaussian i.i.d. with 0 mean and $\sigma^2 = 1$. An indication of the looseness of the bound given by Theorem 3 can be obtained by evaluating both the analytical bound given by (5) as well as the looser one given by (6). We evaluated the pth-order bound for $p = 2$ and $p = 4$ for Gaussian data and only for $n = 2$. For the second-order bound, we obtained 0.5 (analytic) versus 0.25 (looser), whereas for the fourth-order bound, we got 0.43 versus 0.1. For larger values of $n$, however, the numerical evaluation of the analytic bound becomes an impractical task of formidable complexity due to the required multiple integration over the $n$-dimensional data space.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments on the original manuscript.

REFERENCES

Reduced-Rank Adaptive Filtering
J. Scott Goldstein and Irving S. Reed

Abstract—A novel rank reduction scheme is introduced for adaptive filtering problems. This rank reduction method uses a cross-spectral metric to select the optimal lower dimensional subspace for reduced-rank adaptive filtering as a function of the basis vectors of the full-rank space.

I. INTRODUCTION

This correspondence is concerned with rank reduction in adaptive signal processing. The goal of reduced-rank adaptive filtering is to find a lower dimensional filter that yields a steady-state performance that is as close as possible to that obtained by the full-rank solution. The motivation for rank reduction can be attributed to many factors. First, it is very common for the problem under consideration to be overmodeled. In this case, the rank may be reduced to the dimension of the signal subspace to suppress the noise. Second, it could be required that the adaptive filter be of a particular order, perhaps lower than the dimension of the signal subspace, due to complexity constraints or other real-time implementation requirements. For this compression problem, it is desired that the steady-state performance of the reduced-rank filter be as close as possible to the full-rank optimal solution for each value of the filter rank. Clearly, the solution to the compression problem satisfies the overmodeling problem when the rank of the filter equals the dimension of the signal subspace. Finally, the popular least squares (LS) class of algorithms converge as a function of the filter order, implying that lower-rank filters converge faster.

Previous work in reduced-rank adaptive filtering has been concerned primarily with the overmodeling problem [1]–[8]. For notational purposes, the full-rank problem is defined to be of dimension \( N \). In addition, let \( D \) denote the dimension of the signal subspace. With this notation, the previous work on rank reduction consisted of an estimation of the covariance matrix of the observed data and then a determination of its singular value decomposition (SVD). Those eigenvectors corresponding to the largest \( D \) singular values are then retained to form the rank \( D \) eigensubspace in which the reduced-rank filter will operate. This method is very effective if the proper dimension \( D \) is known exactly. In the event that this dimension is not known, then one must either estimate it or choose a rank large enough to ensure that at least \( D \) eigenvectors are retained. If few eigenvectors are retained, the performance will suffer greatly.

In this correspondence, a metric is found that relates directly to the data space and provides a measure of the cross-spectral energy projected along each basis vector. Those \( M \) bases for which this energy contribution is greatest are retained. It is demonstrated that this cross-spectral metric obtains the best low-rank filter as a function of the basis used. In addition, for the overmodeling problem, this metric provides a more robust criterion than the largest eigenvalue criteria for \( M < D \). This counterintuitive result yields a steady-state solution that is the upper bound on the performance of an adaptive filter that operates in the rank \( M \) eigensubspace for all \( M \leq N \).

II. THE FULL-RANK LS PROBLEM

Let \( X \) denote an \( L \times N \) data observation matrix, and let \( d \) be some desired data vector of dimension \( L \). The goal of the LS problem is to find the best approximation of \( d \) that is solely a weighted linear combination of the \( N \) column vectors that compose \( X \). The error \( e \) to be minimized is given by

\[
e = d - Xw
\]

where \( w \) is the \( N \)-dimensional weight vector to be determined.

The LS method estimates the \( N \times N \) covariance matrix \( R_x = X^H X \) and the \( N \times 1 \) cross-correlation vector between the observed data and the desired signal vector \( r_{x,d} = X^H d \). The standard LS solution for \( w \) is then provided by \( w_{LS} \), which is computed as

\[
w_{LS} = R_x^{-1} r_{x,d}.
\]

This solution yields \( e_{LS} = d - X w_{LS} \) as the error vector with minimum Euclidean norm. The error \( e_{LS} \) is orthogonal to the column space of \( X \).

III. THE REDUCED-RANK LS PROBLEM

The SVD of the data matrix \( X \) is obtained next as follows:

\[
X = USV^H
\]

where \( U \) \( L \times N \) orthonormal matrix of left singular vectors
\( S \) \( N \times N \) diagonal matrix of singular values
\( V \) \( N \times N \) orthonormal matrix of right singular vectors of \( X \)
(e.g., see [9]).