On the Equivalence Relationships among Fisher Information, Shannon Measures and Variance

Yuan Xinjie, Tianren Peng, Zhenyu Liu, Shao-Lun Huang*
Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China
{yuanxj23, ptr22}@mails.tsinghua.edu.cn, {shaolun.huang, zhenyuliu}@sz.tsinghua.edu.cn

***Abstract*—This paper studies the fundamental relationships between mutual information, parametric and non-parametric Fisher information, minimum mean squared error (MMSE), and conditional variance. In particular, we introduce a unified and novel framework called Equivalent Polynomial Representation (EPR), which represents these quantities within a common polynomial space for dual-scaled parametric MIMO Gaussian channels. Within this framework, classical identities such as de Bruijn's identity and the I-MMSE relationship, are recovered as special cases. Beyond these, the EPR framework yields new fourth-degree identities connecting MMSE derivatives, parametric Fisher information, and conditional variance under general parametric scaling regimes. These results reveal hidden algebraic structure underlying the interplay between estimation-theoretic and information-theoretic quantities.**

***Index Terms*—de Bruijn's identity, I-MMSE, Fisher information, parametric estimation**

I. Introduction

Information theory and statistical estimation theory are closely connected through fundamental identities relating entropy, Fisher information, and estimation error. Two landmark results exemplify this connection: de Bruijn's identity [1] and the I-MMSE (mutual information-minimum mean squared error) relationship [2].

De Bruijn's identity states that for the additive Gaussian noise channel $Y = X + \sqrt{t}N$ where $N \sim \mathcal{N}(0,1)$ is independent of X, the differential entropy satisfies

$$\frac{\partial}{\partial t} h(Y) = \frac{1}{2} J(Y),$$

where $J(Y)$ denotes the non-parametric Fisher information (NPFI). Subsequently, Guo, Shamai, and Verdú [2] established the I-MMSE relationship for the scaled signal channel $Y = \sqrt{t}X + N$:

$$\frac{\partial}{\partial t} I(X;Y) = \frac{1}{2}\,\mathrm{MMSE}(t),$$

where $I(X;Y)$ denotes mutual information (MI) and $\mathrm{MMSE}(t)$ denotes the mean squared estimation error of X from Y. These identities have inspired extensive research connecting information-theoretic quantities and estimation error, with applications spanning communication systems, signal processing, and statistical inference [3]–[9].

However, existing works predominantly focus on NPFI, which measures sensitivity with respect to observations. In contrast, parametric Fisher information (PFI), which quantifies sensitivity with respect to distributional parameters, is more relevant for sensing-oriented settings where communication and parameter estimation are both of interest [10]–[15]. The distinction is fundamental: NPFI involves spatial derivatives, whereas PFI involves parametric derivatives.

For PFI, the celebrated Cramér-Rao bound [16] provides a fundamental inequality relating PFI to the variance of parameter estimators:

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{J_\theta},$$

where J_θ is the parametric Fisher information for parameter θ. However, this inequality concerns parameter estimation accuracy, which is conceptually distinct from signal estimation error in the I-MMSE relationship. More importantly, systematic relationships connecting PFI, MI, MMSE, and conditional variance remain largely unexplored.

In this paper, we address this gap by proposing a novel unified framework, called Equivalent Polynomial Representation (EPR), for dual-scaled MIMO Gaussian channels where both signal and noise intensities vary parametrically. The key insight is that MI, PFI, NPFI, MMSE, and conditional variance can be represented in a common polynomial space under mild boundary conditions, enabling direct algebraic comparison across quantities. Within this framework, classical identities—de Bruijn's identity and the I-MMSE relationship—are recovered as special cases. More importantly, we derive three new fourth-degree identities: a new MMSE-derivative/conditional-covariance relationship and two new PFI-variance relationships under special scaling laws. These results reveal hidden connections between estimation and communication, and provide analytical insights for parameter-aware communication and sensing design.

II. Channel Model and Preliminaries

We consider the following dual-scaled MIMO Gaussian channel model:

$$Y = r(t)\, H\, X + r_N(t) Z, \tag{1}$$

where $Y \in \mathbb{R}^L$ is the received signal, $H \in \mathbb{R}^{L\times K}$ is a random but fixed channel matrix, $X \in \mathbb{R}^K$ is the transmitted signal, and $Z \in \mathbb{R}^L$ is the additive Gaussian noise following $\mathcal{N}(0, I_L)$, independent of X. I_L denotes the L-dimensional identity matrix. The functions $r(t)$ and $r_N(t)$ are differentiable

Corresponding author: Shao-Lun Huang, whose research is supported in part by National Key R&D Program of China under Grant 2021YFA0715202, the National Natural Science Foundation of China under Grants 62571296.

scalar functions with parameter t being an abstract scalar parameter encompassing SNR, noise variance, or sensing parameters as special cases. $r(t)$ and $r_N(t)$ capture a wide range of practical scenarios such as range-dependent attenuation in radar or sensing settings (where both signal strength and noise level vary with distance, e.g., $r(t) \propto t^{-2}, r_N(t) \propto t^{-\alpha}$) and adaptive systems with power control and time-varying interference (where $r(t)$ models transmit power adaptation and $r_N(t)$ captures interference dynamics).

For brevity, we omit t and write r, r', r_N, r'_N for $r(t), r'(t), r_N(t), r'_N(t)$ when clear from context, where the prime denotes derivative w.r.t. t. Probability density functions (PDFs) are denoted by $f(\cdot)$ with appropriate subscripts indicating the associated random variables. For instance, the input X follows an arbitrary distribution with PDF $f_X(x)$.

The MMSE estimator of X given Y is defined as

$$\hat{X}(Y) \triangleq \mathbb{E}\left[X|Y\right]. \tag{2}$$

We define the conditional covariance matrix of the transformed signal HX

$$\mathrm{Cov}\left(HX|Y;t\right) \triangleq \mathbb{E}\left[\left(HX - H\hat{X}(Y)\right)\left(HX - H\hat{X}(Y)\right)^{\mathrm{T}}\Big|Y\right]. \tag{3}$$

The error covariance matrix of $H\hat{X}$ is defined as the expectation of the conditional covariance:

$$\mathrm{Cov}_{H\hat{X}}(t) \triangleq \mathbb{E}\left[\mathrm{Cov}\left(HX|Y;t\right)\right]. \tag{4}$$

The associated MMSE is defined as the trace of the error covariance matrix:

$$\mathrm{MMSE}_{HX}(t) \triangleq \mathrm{Tr}\left(\mathrm{Cov}_{H\hat{X}}(t)\right), \tag{5}$$

where $\mathrm{Tr}(\cdot)$ denotes the matrix trace operator. The differential entropy of Y is defined as

$$h(Y;t) \triangleq -\mathbb{E}\left[\log f_Y(Y;t)\right]. \tag{6}$$

The mutual information between X and Y is given by

$$I(X;Y;t) \triangleq h(Y;t) - h(Y|X;t) \tag{7}$$

which reduces to $h(Y;t) - \frac{L}{2}\log\left(2\pi e r_N^2\right)$ under Gaussian noise. The non-parametric Fisher information matrix measures the sensitivity of the log-likelihood with respect to observations:

$$J(Y;t) \triangleq \mathbb{E}\left[\left(\nabla_y \log f_Y(Y;t)\right)\left(\nabla_y \log f_Y(Y;t)\right)^{\mathrm{T}}\right], \tag{8}$$

where ∇_y denotes the gradient operator with respect to y. Equivalently, this quantity can be interpreted as Fisher information with respect to a virtual location parameter that shifts the observation vector. In contrast, the parametric Fisher information quantifies the sensitivity of the log-likelihood with respect to the parameter t:

$$J_Y(t) \triangleq \mathbb{E}\left[\left(\frac{\partial}{\partial t}\log f_Y(Y;t)\right)^2\right]. \tag{9}$$

Note that non-parametric Fisher information involves spatial derivatives with respect to observations, whereas parametric Fisher information involves derivatives with respect to the parametric scaling variable t.

III. Equivalent Polynomial Representation Framework

In this section, we develop a unified analytical framework based on the novel concept of Equivalent Polynomial Representation. We first define EPR and establish its fundamental equivalence properties, and then derive an order-reduction lemma that enables systematic simplification of higher-order representations. Unless otherwise specified, all integrals are taken over $\mathbb{R}$ or appropriate product spaces, with limits understood to be $\pm\infty$.

A. Equivalent Polynomial Representation

Definition 1 (Equivalent Polynomial Representation)**.** *Let $q(y, x_1, \ldots, x_k)$ be a polynomial function of variables $y \in \mathbb{R}^L$ and $x_1, \ldots, x_k \in \mathbb{R}^K$. If a function $g(t)$ can be represented as*

$$g(t) = \int q(y, x_1, \ldots, x_k) f_Y(y;t) \prod_{i=1}^{k} f_{X|Y}(x_i|y;t)\mathrm{d}x_i\mathrm{d}y,$$

then $q(y, x_1, \ldots, x_k)$ is called an equivalent polynomial representation *of $g(t)$, denoted by*

$$g(t) \overset{\mathscr{I}}{\leftrightarrow} q(y, x_1, \ldots, x_k).$$

Furthermore, if $q_i(y, x_1, \ldots, x_{k_i}), i = 1, 2$ are both EPRs of $g(t)$, they are said to be equivalent, denoted by [1]

$$q_1(y, x_1, \ldots, x_{k_1}) \leftrightarrow q_2(y, x_1, \ldots, x_{k_2}).$$

We further define the order *of an EPR as the minimum number of distinct replicas of x involved in the EPR. We also define the* degree *of $g(t)$ as the degree of the polynomial q with respect to y and x_i's.*

Remark 1. *According to Definition 1, $g(t)$ can be written as a conditional expectation only when $X_1, \ldots, X_k$ are conditionally i.i.d. given Y and each follows $f_{X|Y}(\cdot|Y;t)$. Under this condition,*

$$g(t) = \mathbb{E}\Big[\mathbb{E}\big[q(Y, X_1, \ldots, X_k) \mid Y;t\big]\Big].$$

This gives a probabilistic interpretation of EPR and links it to standard conditional expectation notation.

Taking $\mathrm{Cov}_{H\hat{X}}(t)$ as an example to illustrate the concept of EPR, we first derive its EPR as follows:

$$\begin{aligned}
\mathrm{Cov}_{H\hat{X}}(t) =& \mathbb{E}\left[\mathbb{E}\left[\left(HX - H\hat{X}(Y)\right)\left(HX - H\hat{X}(Y)\right)^{\mathrm{T}}\Big|Y\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[HX(HX)^{\mathrm{T}} \mid Y\right] - \mathbb{E}\left[HX \mid Y\right]\mathbb{E}\left[HX \mid Y\right]^{\mathrm{T}}\right] \\
=& \int \left(Hx_1x_1^{\mathrm{T}}H^{\mathrm{T}} - Hx_1x_2^{\mathrm{T}}H^{\mathrm{T}}\right) \\
& \cdot f_Y(y;t)\prod_{i=1}^{2} f_{X|Y}(x_i|y;t)\mathrm{d}x_i\mathrm{d}y \\
\overset{\mathscr{I}}{\leftrightarrow}& Hx_1x_1^{\mathrm{T}}H^{\mathrm{T}} - Hx_1x_2^{\mathrm{T}}H^{\mathrm{T}}.
\end{aligned} \tag{10}$$

[1]Here, $\overset{\mathscr{I}}{\leftrightarrow}$ denotes the representation relation between a function of parameter t and a polynomial, while $\leftrightarrow$ denotes equivalence between polynomial representations. This relation induces a map $\mathcal{R}$ from admissible polynomials to functions of t, i.e., $g(t) \overset{\mathscr{I}}{\leftrightarrow} q \iff g(t) = \mathcal{R}[q]$. Accordingly, $q_1 \leftrightarrow q_2 \iff \mathcal{R}[q_1] = \mathcal{R}[q_2]$. Since the polynomial arguments may be cumbersome, we use arrow-style relation symbols in the main text instead of explicit map notation.

Since $\mathbb{E}\left[\mathbb{E}\left[HX|Y\right]\mathbb{E}\left[HX|Y\right]^{\mathrm{T}}\right]$ includes two conditional expectations, two replicas x_1 and x_2 are required in the EPR. Therefore, the order of the EPR is 2. The degree of $\mathrm{Cov}_{H\hat{X}}(t)$ is 2 as well, since the polynomial is quadratic in x_i.

Similarly, the non-parametric Fisher information $J(Y;t)$ can also be expressed in EPR form[2]:

$$J\left(Y;t\right) \overset{\mathscr{I}}{\leftrightarrow} \frac{1}{r_N^4}\left(yy^{\mathrm{T}} - ryx_1^{\mathrm{T}}H^{\mathrm{T}} - rHx_1y^{\mathrm{T}} + r^2Hx_1x_2^{\mathrm{T}}H^{\mathrm{T}}\right),$$

which is also of order 2 and degree 2. According to the basic properties of conditional expectations, we can further reduce monomials involving y of order 1 or 0 to eliminate the dependence on y using the following property.

Property 1. *The following equivalence relationships for second-degree polynomials hold:*

1) $yy^{\mathrm{T}} \leftrightarrow r^2Hx_1x_1^{\mathrm{T}}H^{\mathrm{T}} + r_N^2\, I_L$,
2) $yx_1^{\mathrm{T}}H^{\mathrm{T}} \leftrightarrow r\, Hx_1x_1^{\mathrm{T}}H^{\mathrm{T}}$.

Taking trace on both sides preserves the equivalence.

Proof. Expanding into integrals of $\mathbb{E}\left[YY^{\mathrm{T}}\right] = r^2H\mathbb{E}\left[XX^{\mathrm{T}}\right]H^{\mathrm{T}} + r_N^2I_L$ and $\mathbb{E}\left[YX^{\mathrm{T}}H^{\mathrm{T}}\right] = rH\mathbb{E}\left[XX^{\mathrm{T}}\right]H^{\mathrm{T}}$, the equivalence relationships follow directly. □

Appendix provides fourth-degree equivalence relationships by Property 2. Applying Property 1, we obtain NPFI's EPR as

$$J\left(Y;t\right) \overset{\mathscr{I}}{\leftrightarrow} \frac{1}{r_N^2}I_L - \frac{r^2}{r_N^4}Hx_1x_1^{\mathrm{T}}H^{\mathrm{T}} + \frac{r^2}{r_N^4}Hx_1x_2^{\mathrm{T}}H^{\mathrm{T}}, \quad (11)$$

which shares the same second-degree components as the error covariance matrix's EPR, i.e., $x_1x_1^{\mathrm{T}}$ and $x_1x_2^{\mathrm{T}}$.

A crucial property of the EPR framework is that the equivalence relationship "$\leftrightarrow$" preserves linearity.

Proposition 1 (Linearity of EPR Equivalence)**.** *Suppose that for a collection of functions $\{g_i(t)\}_{i=1}^n$, we have $g_i(t) \overset{\mathscr{I}}{\leftrightarrow} q_i(y, x_1, \ldots, x_{k_i})$. If there exist coefficients $\{c_i(t)\}_{i=1}^n$ and a term $c_0(t)$ such that the polynomial representations satisfy*

$$\sum_{i=1}^{n} c_i(t)q_i(y, x_1, \ldots, x_{k_i}) \leftrightarrow c_0(t),$$

then it follows that

$$\sum_{i=1}^{n} c_i(t)g_i(t) = c_0(t).$$

This means that any linear identity established at the polynomial level directly induces a corresponding identity among the associated information-theoretic quantities. In particular, this proposition is the formal bridge that justifies turning coefficient-level EPR equivalences into function-level equalities throughout the paper. As an example, comparing the EPRs in Eq. (10) and Eq. (11), the following proposition holds, connecting NPFI and the error covariance matrix.

Proposition 2. *Consider the dual-scaled MIMO Gaussian channel Eq. (1). The following relationship holds:*

$$J\left(Y;t\right) = -\frac{r^2}{r_N^4}\mathrm{Cov}_{H\hat{X}}(t) + \frac{1}{r_N^2}I_L, \quad (12)$$

or equivalently taking trace on both sides,

$$\mathrm{Tr}\left(J\left(Y;t\right)\right) = -\frac{r^2}{r_N^4}\,\mathrm{MMSE}_{HX}(t) + \frac{L}{r_N^2}. \quad (13)$$

B. Order Reduction Lemma

It is hard to compare EPRs of different orders directly. To address this challenge, an order-reduction lemma is developed, which can reduce higher-order EPRs to lower-order ones. A mild regularity condition is required.

Assumption 1 (Boundary-Term Vanishing Condition)**.** *For a k-th order polynomial $q(y, x_1, \ldots, x_k)$, we define the following auxiliary function:*

$$\phi_q(y) \triangleq \int q\left(y, x_1, \ldots, x_k\right)\prod_{i=1}^{k} f_{XY}(x_i, y; t)\prod_{i=1}^{k}\mathrm{d}x_i. \quad (14)$$

We assume that the following boundary integrals vanish depending on the order of q:

$$\int_{\partial\Omega} \log f_Y(y;t)\phi_q(y)^{\mathrm{T}}n(y)\mathrm{d}S = 0, \qquad k = 1, \quad (15)$$

$$\int_{\partial\Omega} f_Y^{-k+1}(y;t)\phi_q(y)^{\mathrm{T}}\, n(y)\mathrm{d}S = 0, \qquad k \geq 2, \quad (16)$$

where $\Omega = \mathbb{R}^L$, $\partial\Omega$ denotes its boundary at infinity, $n(y)$ is the outward unit normal vector, and $\mathrm{d}S$ is the surface measure.

Remark 2. *Intuitively, Assumption 1 requires that the output distribution's tail decays fast enough to dominate the polynomial growth of q. This condition is satisfied by a broad class of input distributions, including compact-support distributions, Gaussian distributions, and many exponential family distributions (e.g., Laplace, exponential, Gamma).*

Lemma 1 (Order Reduction Lemma)**.** *Let $k \geq 2$ and $q(y, x_1, \ldots, x_k) \in \mathbb{R}^L$ be a polynomial function of order k that satisfies Assumption 1. Then the following $(k{+}1)$-th order EPR can be reduced to k-th order by*

$$r\left(Hx_{k+1}\right)^{\mathrm{T}}q \leftrightarrow \frac{1}{k-1}\Big(-y^{\mathrm{T}}q + \sum_{i=1}^{k}\left(rHx_i\right)^{\mathrm{T}}q + r_N^2\,\nabla_y \cdot q\Big), \quad (17)$$

where $\nabla_y \cdot q$ denotes the divergence of q with respect to y.

Proof. Using auxiliary function $\phi_q(y)$ defined in Eq. (14), and applying the divergence theorem, we can derive the following integration by parts formula:

$$\int f_Y^{-k+1}(y;t)\nabla_y \cdot \phi_q(y)\mathrm{d}y$$
$$= \int_{\partial\Omega} f_Y^{-k+1}(y;t)\phi_q(y)^{\mathrm{T}}n(y)\mathrm{d}S - \int \phi_q^{\mathrm{T}}(y)\nabla_y f_Y^{-k+1}(y;t)\mathrm{d}y.$$

By Assumption 1, the boundary term vanishes. Substituting $\nabla_y f_Y(y;t) = \int \frac{rHx-y}{r_N^2}f_{XY}(x,y;t)\mathrm{d}x$ and simplifying, the first integral can be expressed as

$$\mathrm{LHS} \overset{\mathscr{I}}{\leftrightarrow} -\frac{1}{r_N^2}\sum_{i=1}^{k}\left(y - rHx_i\right)^{\mathrm{T}}q + \nabla_y \cdot q,$$

while the right-hand side can be expressed as

$$\mathrm{RHS} \overset{\mathscr{I}}{\leftrightarrow} -\left(k-1\right)\frac{1}{r_N^2}\left(y - rHx_{k+1}\right)^{\mathrm{T}}q.$$

Equating and rearranging yields the desired result. □

With the order reduction lemma, higher-order EPRs can be reduced eventually to second-order ones by recursively

[2] According to the Gaussian noise setting, we have $f_{Y|X}(y|x;t) = (2\pi r_N^2)^{-L/2}\exp\left(-\|y - rHx\|^2/2r_N^2\right)$, and thus $\nabla_y \log f_{Y|X}(y|x;t) = (rHx - y)/r_N^2$. Further we get $\nabla_y f_Y(y;t) = \int \frac{rHx-y}{r_N^2}f_{XY}(x,y;t)\mathrm{d}x$, and $\nabla_y \log f_Y(y;t) = \int \frac{rHx-y}{r_N^2}f_{X|Y}(x|y;t)\mathrm{d}y$.

applying Lemma 1. The following corollary provides an example of reducing an EPR of order 3 to order 2.

Corollary 1. *The following equivalence holds:*

$$r\left\|Hx_1\right\|^2\left(Hx_2\right)^{\mathrm{T}}\left(Hx_3\right) \leftrightarrow r\left\|Hx_1\right\|^2\left(Hx_1\right)^{\mathrm{T}}\left(Hx_2\right) + r\left\|Hx_1\right\|^2\left\|Hx_2\right\|^2 - \left\|Hx_1\right\|^2 y^{\mathrm{T}} Hx_2.$$

Proof. Invoking Lemma 1 with $q = \left\|Hx_1\right\|^2 Hx_2$ and $k = 2$ yields the desired result. □

IV. Main Results

A wide range of information-theoretic quantities can be expressed in terms of EPRs and finally reduced to second-order EPRs. Such a framework enables direct comparisons and explicit relationships among seemingly disparate quantities.

A. Second-Degree Relationships

Proposition 3. *Consider the dual-scaled MIMO Gaussian channel Eq. (1). Under Assumption 1, the following relationships hold:*

(i) ***Extended I-MMSE Relationship:***

$$\frac{\partial}{\partial t} I(X;Y;t) = \frac{r}{r_N^3}\left(r' r_N - r r_N'\right) \mathrm{MMSE}_{HX}(t). \tag{23}$$

(ii) ***Extended de Bruijn's Identity:***

$$\frac{\partial}{\partial t} h(Y;t) = \frac{r_N}{r}(r r_N' - r' r_N)\mathrm{Tr}\left(J(Y;t)\right) + \frac{r'}{r} L. \tag{24}$$

Proof. The EPR of $\frac{\partial}{\partial t} h(Y;t)$ can be derived as follows:

$$\begin{aligned}
\frac{\partial}{\partial t} h\left(Y;t\right) &\overset{\text{(a)}}{=} -\int \frac{\partial}{\partial t} f_Y(y;t) \log f_Y(y;t) \mathrm{d}y \\
&\overset{\text{(b)}}{=} -\int \left(\nabla_y \cdot \phi_{\tilde{q}}\left(y;t\right)\right) \log f_Y\left(y;t\right) \mathrm{d}y \\
&\overset{\text{(c)}}{=} \int \phi_{\tilde{q}}\left(y;t\right)^{\mathrm{T}} \nabla_y \log f_Y\left(y;t\right) \mathrm{d}y \\
&\quad - \int_{\partial\Omega} \log f_Y\left(y;t\right) \phi_{\tilde{q}}\left(y;t\right)^{\mathrm{T}} n(y) \mathrm{d}S \\
&\overset{\text{(d)}}{=} \int \frac{1}{f_Y(y;t)} \left(\int \tilde{q}(y,x) f_{XY}(x,y;t) \mathrm{d}x\right)^{\mathrm{T}} \\
&\qquad \left(\int \frac{rHx - y}{r_N^2} f_{XY}(x,y;t) \mathrm{d}x\right) \mathrm{d}y \\
&\overset{\mathscr{I}}{\leftrightarrow} \tilde{q}(y,x_1)^{\mathrm{T}} \frac{rHx_2 - y}{r_N^2},
\end{aligned}$$

where step (a) follows from the regularity of $f_Y(y;t)$ that $\int \frac{\partial}{\partial t} f_Y(y;t)\mathrm{d}y = \frac{\partial}{\partial t}\int f_Y(y;t)\mathrm{d}y = 0$, (b) expresses the derivative w.r.t. t as a spatial divergence via the polynomial

$$\tilde{q}(y,x) \triangleq \frac{1}{r_N}\left(-r_N' y + (r r_N' - r' r_N) Hx\right) \tag{25}$$

and its induced function $\phi_{\tilde{q}}(y;t) \triangleq \int f_{XY}(x,y;t)\tilde{q}(y,x)\mathrm{d}x$, such that $\nabla_y \cdot \phi_{\tilde{q}}(y;t) = \frac{\partial}{\partial t} f_Y(y;t)$, (c) applies integration by parts, and (d) invokes Assumption 1 to eliminate the boundary term. The resulting EPR is given by

$$\frac{\partial}{\partial t} h\left(Y;t\right) \overset{\mathscr{I}}{\leftrightarrow} \frac{(r r_N' - r' r_N) r}{r_N^3}\left(x_1^{\mathrm{T}} H^{\mathrm{T}} H x_2 - \left\|Hx_1\right\|^2\right) + \frac{r_N'}{r_N} L. \tag{26}$$

Finally, by Proposition 1, comparing Eq. (10) and Eq. (26) with the fact that $\frac{\partial}{\partial t} I(X;Y;t) = \frac{\partial}{\partial t} h(y;t) - L\frac{r_N'}{r_N}$ yields (i). Likewise, comparing Eq. (11) and Eq. (26) yields (ii). □

Remark 3. *The classical I-MMSE relationship and de Bruijn's identity are recovered as special instances of Proposition 3. Specifically, setting $r(t) = \sqrt{t}, r_N(t) = 1$, (i) reduces to the classical I-MMSE relationship $\frac{\partial}{\partial t} I(X;Y;t) = \frac{1}{2}\mathrm{MMSE}_{HX}(t)$. When $r(t) = 1, r_N(t) = \sqrt{t}$, (ii) recovers de Bruijn's identity $\frac{\partial}{\partial t} h(Y;t) = \frac{1}{2}\mathrm{Tr}\left(J(Y;t)\right)$.*

The proof reveals that all second-degree quantities share a common monomial basis: $x_1 x_1^{\mathrm{T}}$ and $x_1 x_2^{\mathrm{T}}$ (or their traces for scalar forms). These monomials span a two-dimensional space. Since the coefficients of our quantities of interest under this basis are proportional, mutual expressibility is guaranteed. The following corollary establishes a direct connection among the three key quantities without zero-degree terms.

Corollary 2. *Under the same conditions as in Proposition 3, the following identity holds:*

$$\frac{\partial}{\partial t} h(Y;t) = r_N r_N' \mathrm{Tr}\big(J(Y;t)\big) + \frac{r r'}{r_N^2}\,\mathrm{MMSE}_{HX}(t). \tag{27}$$

B. Fourth-Degree Relationships

We present several key novel fourth-degree identities, with proofs deferred to the appendix for brevity.

Theorem 1. *Consider the dual-scaled MIMO Gaussian channel Eq. (1). Under Assumption 1, the following identity holds:*

$$\begin{aligned}
& r_N^3 \frac{\partial}{\partial t} \mathrm{MMSE}_{HX}(t) \\
=& 2\left(r r_N' - r' r_N\right) r \mathbb{E}\left[\mathrm{Tr}\left(\mathrm{Cov}^2\left(HX|Y;t\right)\right)\right].
\end{aligned} \tag{28}$$

This theorem shows that the rate of change of MMSE is "parallel" to the expected trace of the squared conditional covariance, scaled by the difference in signal and noise scaling rates. This theorem also allows us to further relate $\mathbb{E}\left[\mathrm{Tr}\left(\mathrm{Cov}^2\left(HX|Y;t\right)\right)\right]$ with the derivative of non-parametric Fisher information and the second derivative of mutual information using Proposition 3, which we omit here for brevity. The relationship between PFI and variance is further characterized in two special parametric settings.

Theorem 2 (PFI-Variance Relationship: Noise Scaling Only)**.** *Consider the dual-scaled MIMO Gaussian channel Eq. (1) with constant signal scaling $r(t) = c > 0$, under Assumption 1, the PFI satisfies*

$$\begin{aligned}
r_N^4 J_Y(t) =& c^2 r_N r_N' \mathbb{E}\left[\frac{\partial}{\partial t}\mathrm{Tr}\left(\mathrm{Cov}\left(HX|Y;t\right)\right)\right] \\
& - 4c^2\left(r_N'\right)^2 \mathrm{MMSE}_{HX}(t) + 2L r_N^2\left(r_N'\right)^2.
\end{aligned} \tag{29}$$

Theorem 3 (PFI-Variance Relationship: Quadratic Signal-Noise Coupling)**.** *Consider the dual-scaled MIMO Gaussian channel Eq. (1) with $r(t) = c r_N^2(t)$ for some constant $c > 0$, under Assumption 1, the PFI satisfies*

$$\begin{aligned}
r^2 J_Y(t) =& -\frac{1}{4} c^2 r^2 \left(r'\right)^2 \mathbb{E}\left[\mathrm{Var}\left(\left\|HX\right\|^2 |Y;t\right)\right] \\
& + c r\left(r'\right)^2 \mathbb{E}\left[\left\|HX\right\|^2\right] + \frac{1}{2} L\left(r'\right)^2.
\end{aligned} \tag{30}$$

The following corollary illustrates two specific instances of Theorems 2 and 3.

Corollary 3. *If $r(t) = 1, r_N(t) = 1/\sqrt{2t}$, then*

$$J_Y(t) = -\mathbb{E}\left[\frac{\partial}{\partial t}\mathrm{Tr}\left(\mathrm{Cov}\left(HX|Y;t\right)\right)\right] - \frac{2}{t}\mathrm{MMSE}_{HX}(t) + \frac{L}{2t^2}.$$

If $r(t) = t, r_N(t) = \sqrt{t/2}$, then

$$J_Y(t) = -\mathbb{E}\left[\mathrm{Var}\left(\left\|HX\right\|^2 |Y;t\right)\right] + \frac{2}{t}\mathbb{E}\left[\left\|HX\right\|^2\right] + \frac{L}{2t^2}.$$

$$
\begin{aligned}
J_Y(t) \overset{\mathscr{I}}{\leftrightarrow} & \frac{1}{r^2 r_N^6}\big(-a^2 r^4 \|Hx_1\|^4 + (a-b)^2 r^4 \|Hx_1\|^2 \|Hx_2\|^2 - 2(2a-b)(a-b)r^3 \|Hx_1\|^2 y^{\mathrm{T}} Hx_2 \\
& + (2a-b)^2 r^2 (y^{\mathrm{T}} Hx_1)(y^{\mathrm{T}} Hx_2) - 4a(a-b) r_N^2 r^2 \|Hx_1\|^2 + 2L r_N^4 a^2\big) && (18)\\
\frac{\partial}{\partial t}\,\mathrm{MMSE}_{HX}(t) \overset{\mathscr{I}}{\leftrightarrow} & \frac{1}{r^3 r_N^3}\big[-(a-b) r^4 \|Hx_1\|^2 (Hx_1)^{\mathrm{T}} (Hx_2) + (a-b) r^4 \|Hx_1\|^2 \|Hx_2\|^2 \\
& -(a-b) r^3 \|Hx_1\|^2 y^{\mathrm{T}} Hx_2 + (a-b) r^2 (Hx_1)^{\mathrm{T}} (Hx_2)\, y^{\mathrm{T}} y - L(a-b) r_N^2 r^2 (Hx_1)^{\mathrm{T}} (Hx_2)\big] && (19)\\
\mathbb{E}\left[\frac{\partial}{\partial t}\mathrm{Tr}\left(\mathrm{Cov}\left(HX|Y;t\right)\right)\right] \overset{\mathscr{I}}{\leftrightarrow} & \frac{1}{r_N^3 r^3}\big[-a r^4 \|Hx_1\|^4 + (a-b) r^4 \|Hx_1\|^2 \|Hx_2\|^2 + (3b-4a) r^3 \|Hx_1\|^2 y^{\mathrm{T}} Hx_2 \\
& - 2(b-2a) r^2 (Hx_1)^{\mathrm{T}} y\, (Hx_2)^{\mathrm{T}} y + 2(b-2a) r^2 r_N^2 (Hx_1)^{\mathrm{T}} (Hx_2)\big] && (20)\\
\mathbb{E}\left[\mathrm{Tr}\left(\mathrm{Cov}^2\left(HX|Y;t\right)\right)\right] \overset{\mathscr{I}}{\leftrightarrow} & \frac{1}{2r^4}\big[-r^4 \|Hx_1\|^2 (Hx_1)^{\mathrm{T}} (Hx_2) + r^4 \|Hx_1\|^2 \|Hx_2\|^2 - r^3 \|Hx_1\|^2 (Hx_2)^{\mathrm{T}} y \\
& + r^2 (Hx_1)^{\mathrm{T}} (Hx_2)\, y^{\mathrm{T}} y - r_N^2 r^2 L\, (Hx_1)^{\mathrm{T}} (Hx_2)\big] && (21)\\
\mathbb{E}\left[\mathrm{Var}\left(\|HX\|^2|Y;t\right)\right] \overset{\mathscr{I}}{\leftrightarrow} & \frac{1}{r^4}\left[r^4 \|Hx_1\|^4 - r^4 \|Hx_1\|^2 \|Hx_2\|^2\right] && (22)
\end{aligned}
$$

These results provide engineering insight for sensing-oriented systems. For example, Theorem 2 shows that maximizing PFI and minimizing MMSE are not conflicting objectives. This shows that sensing and communication can be effectively integrated with $\mathbb{E}\left[\frac{\partial}{\partial t}\mathrm{Tr}\left(\mathrm{Cov}\left(HX|Y;t\right)\right)\right]$ being an indicator of the system capability. In Corollary 3, the two scaling regimes indicate that different terms dominate in different operating ranges with respect to the parameter t, suggesting distinct optimization priorities.

Unlike in second-degree space with only two basis elements, fourth-degree space has higher dimensionality, requiring more than two quantities to form a generally held linear relationship. Relatively simple cases arise when the parametric functions $r(t), r_N(t)$ satisfy certain constraints, as demonstrated in the above two theorems. When reducing the MIMO channel to a scalar channel with $L = K = 1$, $H = 1$, fourth-degree EPRs form a five-dimensional space, with basis elements $x_1^4, x_1^3 x_2, x_1^2 x_2^2, x_1^2 x_2 y, x_1 x_2 y^2$. Consequently, at most five linearly independent fourth-degree quantities can fully characterize the fourth-degree components of PFI. For example, in the scalar case, $\mathbb{E}\left[X^4\right]$, $\mathbb{E}\left[\frac{\partial}{\partial t}\mathrm{Var}\left(X|Y;t\right)\right]$, $\mathbb{E}\left[\mathrm{Var}^2\left(X|Y;t\right)\right]$, $\mathbb{E}\left[\left(X-\mathbb{E}\left[X|Y;t\right]\right)^4\right]$, $\mathbb{E}\left[\mathrm{Var}\left(X^2|Y;t\right)\right]$ form such a set of quantities to describe PFI.

V. Conclusion

This paper proposes a unified equivalent polynomial representation framework for dual-scaled MIMO Gaussian channels, unifying parametric/nonparametric Fisher information, mutual information, minimum mean squared error, and conditional variance. Within this framework, extended de Bruijn's identity and I-MMSE relationship are derived, and novel PFI-variance connections are established. These results deepen the links between information theory and parametric estimation, with potential relevance to sensing-oriented communication problems.

Appendix

Property 2. *The following fourth-degree equivalence relationships hold:*

1) $\left(y^{\mathrm{T}} y\right)^2 \leftrightarrow r^4 \|Hx_1\|^4 + 2(L+2) r^2 r_N^2 \|Hx_1\|^2 + L(L+2) r_N^4$
2) $y^{\mathrm{T}} y y^{\mathrm{T}} Hx_1 \leftrightarrow r^3 \|Hx_1\|^4 + (L+2) r r_N^2 \|Hx_1\|^2$
3) $\|Hx_1\|^2 y^{\mathrm{T}} y \leftrightarrow r^2 \|Hx_1\|^4 + L r_N^2 \|Hx_1\|^2$
4) $\|Hx_1\|^2 (Hx_1)^{\mathrm{T}} y \leftrightarrow r \|Hx_1\|^4$

Proof of Theorem 1, 2, and 3. For notational simplicity, we first define two auxiliary scalars

$$a(t) \triangleq r(t) r_N'(t), \quad b(t) \triangleq r'(t) r_N(t).$$

Then using the equivalence properties Property 1 and 2, and the order reduction lemma Lemma 1, we can obtain second-order EPRs of the quantities involved in these theorems in Eq. (18)–Eq. (22). Detailed derivations are omitted for brevity.

By Proposition 1, once the corresponding EPRs are matched, the associated function-level identities follow. Comparing Eq. (19) and Eq. (21) yields Theorem 1. For Theorem 2, $r(t)$ being constant implies $b(t) = 0$. Then Eq. (18) and Eq. (20) share the same fourth-degree basis elements with proportional coefficients. The MMSE Eq. (10) then accounts for the remaining second-degree terms.

For Theorem 3, $r(t) = c r_N^2(t)$ implies $b(t) = 2a(t)$. Then $J_Y(t)$ remains only two terms of fourth-degree monomials, i.e., $\|Hx_1\|^4$ and $\|Hx_1\|^2 \|Hx_2\|^2$ in Eq. (18), with coefficients differing only in sign. $\mathbb{E}\left[\mathrm{Var}\left(\|HX\|^2|Y;t\right)\right]$ in Eq. (22) then makes up for the fourth-degree structure, while signal power $\mathbb{E}\left[\|X\|^2\right]$ accounts for the remaining second-degree terms. □

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[2] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[3] D. Guo, Y. Wu, S. S. Shitz, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

[4] D. Guo, S. Shamai (Shitz), and S. Verdú, "The interplay between information and estimation measures," *Foundations and Trends® in Signal Processing*, vol. 6, no. 4, pp. 243–429, Nov. 2013.

[5] S. Verdú, "On channel capacity per unit cost," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[6] I. Valero-Toranzo, S. Zozor, and J.-M. Brossier, "Generalization of the de Bruijn identity to general ϕ -entropies and ϕ -Fisher informations," *IEEE Transactions on Information Theory*, vol. 64, no. 10, pp. 6743–6758, Oct. 2018.

[7] O. Johnson and S. Guha, "A de Bruijn identity for discrete random variables," in *2017 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2017, pp. 898–902. [Online]. Available: https://ieeexplore.ieee.org/document/8006658

[8] S. Park, E. Serpedin, and K. Qaraqe, "On the equivalence between Stein and de Bruijn identities," *IEEE Transactions on Information Theory*, vol. 58, no. 12, pp. 7045–7067, Dec. 2012.

[9] J. Riba and F. de Cabrera, "A proof of de Bruijn identity based on generalized Price's theorem," in *2019 IEEE International Symposium on Information Theory (ISIT)*, Jul. 2019, pp. 1–5.

[10] X. Yuan, T. Peng, H. Fu, and S.-L. Huang, "Unbounded multi-letter marginal gain for bistatic integrated sensing and communication," in *2025 Asia Pacific Workshop on Data Science and Information Theory*, Shenzhen, Guangdong, China, Oct. 2025.

[11] K. Wu, J. Pegoraro, F. Meneghello, J. A. Zhang, J. O. Lacruz, J. Widmer, F. Restuccia, M. Rossi, X. Huang, D. Zhang, G. Caire, and Y. J. Guo, "Sensing in bistatic ISAC systems with clock asynchronism: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 41, no. 5, pp. 31–43, Sep. 2024.

[12] X. Yuan, T. Peng, H. Fu, and S.-L. Huang, "Multi-letter analysis of marginal rate-Fisher information tradeoff: Unbounded gain," in *2025 Asia Pacific Workshop on Data Science and Information Theory*, Shenzhen, Guangdong, China, Oct. 2025.

[13] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, "On the fundamental tradeoff of integrated sensing and communications under Gaussian channels," *IEEE Transactions on Information Theory*, vol. 69, no. 9, pp. 5723–5751, Sep. 2023.

[14] T. Jiao, K. Wan, Z. Wei, Y. Geng, Y. Li, Z. Yang, and G. Caire, "Information-theoretic limits of bistatic integrated sensing and communication," Mar. 2025.

[15] H. Joudeh and F. M. J. Willems, "Joint communication and binary state detection," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 1, pp. 113–124, Mar. 2022.

[16] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc*, vol. 37, no. 3, pp. 81–91, 1945.