and upon applying the vector calculus identity

\[ \nabla \cdot (\nabla \psi) = \nabla \cdot \psi + \psi \cdot \nabla \nabla \]

to the first term, (13) follows.

**Corollary:** The divergence of \( s(x) \) equals its variance at \( x \); i.e.,

\[ \nabla \cdot s(x) = -V(x). \]  \( (14) \)

This corollary relates spatial variations in \( s(x) \) to conditional expectations at \( x \); and, in addition, it enables us to state that the CME is completely specified by its variance function \( V(x) \). This statement is a consequence of Helmholtz's theorem, which states that a vector is completely specified by its divergence and curl. (Recall that as \( \nabla \cdot s(x) \) is conservative, \( \nabla \times \nabla s(x) = 0 \).)

The primary value of Property 3 and its corollary is that they relate both the likelihood ratio and CME to a conditional variance function of the signal. Hence, by viewing \( L(x) \) as a potential function and \( s(x) \) as a conservative vector field, the mathematics of potential theory can be applied to solving problems in signal detection theory.

### III. Conclusions

This correspondence has noted a fundamental property relating optimum detection and CME for random signals in white Gaussian noise for discrete-time processes, and has discussed the role of the estimation-correlation operation in forming an optimum decision statistic.

By viewing the log-likelihood ratio as a potential function, the CME of the signal was shown to constitute a conservative vector field. This concept was used to show the intimate connection between spatial variation (divergence) of the CME and the conditional signal variance. These results suggest that the mathematics of potential theory might play an important role in furthering the theory of signal detectability.


---

Note on Optimal Selection of Independent Binary-Valued Features for Pattern Recognition

**Abstract**—Given a set of conditionally independent binary-valued features, a counter example is given to a possible claim that the best subset of features must contain the best single feature.

Recently, Elashoff et al.\(^1\) showed that for optimal selection of a subset of independent binary-valued features, the features generally may not be evaluated independently. Specifically, an example is given\(^1\) in which, given three independent variables \( x_1, x_2, \) and \( x_3 \) such that \( s(x_1) < s(x_2) < s(x_3) \), where \( s(x) \) is the error probability when the \( i \)th variable alone is used, the first and third variables are jointly better than the first and second variables. In other words, \( s(x_1, x_3) < s(x_1, x_2) \).


---

Godfried T. TOUSSAINT


Univ. British Columbia

Vancouver, B.C., Canada

**Comments on "A Modified Figure of Merit for Feature Selection in Pattern Recognition"**

In a recent correspondence\(^{1}\) a modification of the conventional mutual-information effectiveness criterion for feature selection in pattern recognition was described. However, there seems to be some confusion between selecting a subset of features and selecting features individually. This apparent confusion may confuse the reader further.

Manuscript received August 15, 1970; revised January 27, 1971.

---

Manuscript received August 7, 1970; revised January 27, 1971. This research was supported in part by the National Research Council of Canada under Grant NR C A-3308 and by the Defence Research Board of Canada under Grant DGR 2801-30.