Combining Detection with Other Tasks of Information Processing

Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 3200004, Israel

IZS 2016, Zurich, Switzerland, March 2016

Separate vs. Joint Detection and Info Processing

- Classical hypothesis testing: optimal decision – based on the LRT.
- In certain applications, signal detection is only the first phase.
- Second–phase processing: decoding, estimation, compression, ...
- Conventional approach: separate detection & second–phase processing:
  - Apply the LRT regardless of the second–phase task.
  - Apply the optimal second–phase task regardless of the detection.
- Alternative: joint detection and second–phase processing:
  - Optimal decision rule: incorporates cost of the second–phase task.
  - Optimal second phase–task: uses the fact that a signal was detected.

Related Work

- Moustakides (2011), Moustakides *et al.* (2012): parameter estimation.
- Yilmaz *et al.* (2013): sequential estimation.
- Wang (2010), Wang *et al.* (2011): slotted asynchronous communication.
- Weinberger & Merhav (2014): codeword or noise?
- Weinberger & Merhav (2015): channel det. + decoding (Wednesday).
- Merhav (2015): data compression.

Generic Problem Definition

Let $Y \sim P_0/P_1$ be an observable.

Let $\ell : \mathcal{Y} \to \mathbb{R}$ be a cost function associated with the second–phase task.

Find a partition the observation space $\mathcal{X}$ into Ω and Ω^c:

$$\begin{aligned} \text{minimize} \quad & \mathbf{E}_1\{\ell(Y)|Y \in \Omega\} \\ \text{subject to} \quad & P_0(\Omega) \leq \epsilon_{\text{FA}} \\ & P_1(\Omega^c) \leq \epsilon_{\text{MD}} \end{aligned}$$

or, alternatively, replace the FA constraint by a constraint on:

$$\mathbf{E}_0\{\ell(Y)|Y \in \Omega\} \cdot P_0(\Omega)$$

Comment: ϵ_{MD} and ϵ_{FA} cannot be both too small (due to the Neyman–Pearson lemma). Relaxing the tension between them creates more room for minimizing the objective.

Generic Problem Definition (Cont'd)

More formally, find:

$$\begin{aligned} \text{minimize} \quad & \sum_{y \in \Omega} f(y) \\ \text{subject to} \quad & \sum_{y \in \Omega} g(y) \leq G \\ & \sum_{y \in \Omega^c} h(y) \leq H \end{aligned}$$

where the minimization is over all subsets $\{\Omega\}$ and where G and H are prescribed numbers.

A Simple Extension of the Neyman–Pearson Lemma

Let f, g and h be any three functions from $\mathcal{X}$ to $\mathbb{R}$ and let

$$\Omega_\star = \{y : \; f(y) + a \cdot g(y) \leq b \cdot h(y)\},$$

where $a \geq 0$ and $b \geq 0$ are fixed numbers. Let Ω be any other subset of $\mathcal{X}$. If

$$\sum_{y \in \Omega} g(y) \leq \sum_{y \in \Omega_\star} g(y)$$

and

$$\sum_{y \in \Omega^c} h(y) \leq \sum_{y \in \Omega_\star^c} h(y)$$

then

$$\sum_{y \in \Omega_\star} f(y) \leq \sum_{y \in \Omega} f(y).$$

In other words, no competing partition Ω dominates Ω^* in all three criteria.

Comments

$$\Omega_\star = \{y: \ f(y) + a \cdot g(y) \le b \cdot h(y)\}.$$

- a and b – two “Lagrange multipliers” for controlling the two constraints.
- Classic N–P Lemma: special case of $a = 0$ (drops the g–constraint).
- Letting $a, b \to \infty$ with fixed a/b:
  - Full tension between the constraints.
  - Only one Ω satisfies both constraints.
  - No room for optimization.
  - Separate detection and second–phase task.

Application No. 1 – "Codeword or Noise?"

Weinberger & Merhav (2014):
$\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ is the output of a channel $W(\boldsymbol{y}|\boldsymbol{x})$ fed by $\boldsymbol{x} = (x_1, \ldots, x_n)$.

$\mathcal{H}_0: \quad \boldsymbol{Y} \sim W(\boldsymbol{y}|0^n)$ no transmission; $\boldsymbol{Y}$ is pure noise.

$\mathcal{H}_1: \quad \boldsymbol{Y} \sim \frac{1}{M} \sum_{\boldsymbol{x} \in \mathcal{C}} W(\boldsymbol{y}|\boldsymbol{x})$ $\boldsymbol{Y}$ is a noisy version of a message.

Joint detection and decoding:

$$\text{maximize} \quad \overbrace{\frac{1}{M} \sum_{\boldsymbol{y} \in \Omega} \max_{\boldsymbol{x} \in \mathcal{C}} W(\boldsymbol{y}|\boldsymbol{x})}^{P_{\mathsf{C}}}$$

$$\text{subject to} \quad \sum_{\boldsymbol{y} \in \Omega} W(\boldsymbol{y}|0^n) \leq \epsilon_{\mathsf{FA}}$$

$$\sum_{\boldsymbol{y} \in \Omega^c} \frac{1}{M} \sum_{\boldsymbol{x} \in \mathcal{C}} W(\boldsymbol{y}|\boldsymbol{x}) \leq \epsilon_{\mathsf{MD}}$$

“Codeword or Noise” (Cont’d)

Here we can apply the extended N–P lemma with the following assignments:

$$\begin{aligned} f(\boldsymbol{y}) &= -\frac{1}{M} \cdot \max_{\boldsymbol{x} \in \mathcal{C}} W(\boldsymbol{y}|\boldsymbol{x}) \\ g(\boldsymbol{y}) &= W(\boldsymbol{y}|0^n) \\ h(\boldsymbol{y}) &= \frac{1}{M} \sum_{\boldsymbol{x} \in \mathcal{C}} W(\boldsymbol{y}|\boldsymbol{x}) \end{aligned}$$

The resulting detector–decoder:

$$\Omega_\star = \left\{ \boldsymbol{y} : \; a \cdot \sum_{\boldsymbol{x} \in \mathcal{C}} W(\boldsymbol{y}|\boldsymbol{x}) + \max_{\boldsymbol{x} \in \mathcal{C}} W(\boldsymbol{y}|\boldsymbol{x}) \leq b \cdot W(\boldsymbol{y}|0^n) \right\}.$$

For $\boldsymbol{y} \in \Omega_\star$, apply ordinary ML decoding.

- For a and b to affect error exponents, $a = e^{n\alpha}$ and $b = e^{n\beta}$.
- Application no. 2 – channel det.–dec.: $W(\boldsymbol{y}|0^n) \to \frac{1}{M} \sum_{\boldsymbol{x} \in \mathcal{C}} V(\boldsymbol{y}|\boldsymbol{x})$.

Application No. 3 – Lossless Compression

A seemingly natural goal would be to solve the problem:

$$\begin{aligned} \text{minimize} \quad & \mathbf{E}_1\{L(\boldsymbol{Y})|\boldsymbol{Y} \in \Omega\} \\ \text{subject to} \quad & P_0(\Omega) \le \epsilon_{\text{FA}} \\ & P_1(\Omega^c) \le \epsilon_{\text{MD}} \end{aligned}$$

However, in this case, it makes sense to impose exponentially decaying MD and FA probabilities:

$$\begin{aligned} \epsilon_{\text{MD}} &= \exp\{-nE_{\text{MD}}\} \\ \epsilon_{\text{FA}} &= \exp\{-nE_{\text{FA}}\} \end{aligned}$$

in which case, $P_1(\Omega) \approx 1$, and so, $\mathbf{E}_1\{L(\boldsymbol{Y})|\boldsymbol{Y} \in \Omega\} \approx \mathbf{E}_1\{L(\boldsymbol{Y})\}$, and the problem actually decouples into separate detection and compression.

Lossless Compression – Exponential Moments

Consider now

$$\text{minimize} \quad \mathbf{E}_1\{\exp\{\theta L(\boldsymbol{Y})\}|\boldsymbol{Y} \in \Omega\} \qquad \theta > 0$$
$$\text{subject to} \quad P_0(\Omega) \le \epsilon_{\text{FA}}$$
$$P_1(\Omega^c) \le \epsilon_{\text{MD}}$$

For a given Ω,

$$L^*(\boldsymbol{y}) = -\log\left[\frac{[P_1(\boldsymbol{y})]^{1/(1+\theta)}}{\sum_{\boldsymbol{y}'\in\Omega}[P_1(\boldsymbol{y}')]^{1/(1+\theta)}}\right], \qquad \boldsymbol{y} \in \Omega$$

$$\mathbf{E}_1\{\exp[\theta L^*(\boldsymbol{Y})]|\boldsymbol{Y} \in \Omega\} \approx \left(\sum_{\boldsymbol{y}\in\mathcal{Y}^n}[P_1(\boldsymbol{y})]^{1/(1+\theta)}\right)^{1+\theta},$$

thus, we choose

$$f(\boldsymbol{y}) = [P_1(\boldsymbol{y})]^{1/(1+\theta)},\ g(\boldsymbol{y}) = P_0(\boldsymbol{y}),\ h(\boldsymbol{y}) = P_1(\boldsymbol{y}).$$

Lossless Compression – Exponential Moments (Cont'd)

The resulting detector, for $a = e^{n\alpha}, b = e^{n\beta}$:

$$\Omega_\star = \{\boldsymbol{y}: \ [P_1(\boldsymbol{y})]^{1/(1+\theta)} + e^{n\alpha} P_0(\boldsymbol{y}) \leq e^{n\beta} P_1(\boldsymbol{y})\},$$

or, equivalently,

$$\Omega_\star = \left\{ \boldsymbol{y}: \quad \underbrace{\frac{P_1(\boldsymbol{y})}{P_0(\boldsymbol{y})}}_{\text{ordinary LR}} \quad \cdot \underbrace{\left(1 - e^{-n\beta}[P_1(\boldsymbol{y})]^{-\theta/(1+\theta)}\right)}_{\text{correction factor}} \geq e^{n(\alpha-\beta)} \right\}.$$

- Unlike classic LRT, here the test statistics doesn't depend only on the LR.
- Correction factor: reject $\boldsymbol{y}$'s with small $P_1(\boldsymbol{y})$ – cost of coding is high.
- Error exponents can easily be analyzed using the method of types.

Another Variant of the Problem

Consider now the problem:

$$\begin{aligned} &\text{minimize} \quad \mathbf{E}_1\{\exp\{\theta L(\boldsymbol{Y})\}|\boldsymbol{Y} \in \Omega\} \\ &\text{subject to} \quad \mathbf{E}_0\{\exp\{\theta L(\boldsymbol{Y})\}|\boldsymbol{Y} \in \Omega\} \cdot P_0(\Omega) \leq \epsilon_{\text{FA}} \\ &\qquad\qquad P_1(\Omega^c) \leq \epsilon_{\text{MD}} \end{aligned}$$

Here, we would like to use

$$g(\boldsymbol{y}) = P_0(\boldsymbol{y}) \exp\{\theta L^*(\boldsymbol{y})\} = P_0(\boldsymbol{y})[P_1(\boldsymbol{y})]^{-\theta/(1+\theta)} \left(\sum_{\boldsymbol{y}' \in \Omega} [P_1(\boldsymbol{y}')]^{1/(1+\theta)} \right)^{\theta}.$$

Difficulty: The extended N–P lemma cannot be applied since g depends on Ω.

A Possible Remedy

Consider the length function of a universal code $L(\boldsymbol{y}) \approx n\hat{H}(\boldsymbol{y})$, which is asymptotically optimal for every memoryless P and for every θ. Now, consider

$$g(\boldsymbol{y}) = P_0(\boldsymbol{y}) \cdot \exp\{n\theta\hat{H}(\boldsymbol{y})\}.$$

Detection rule:

$$\Omega_\star = \left\{\boldsymbol{y} :\ P_1(\boldsymbol{y}) \exp\{n\theta\hat{H}(\boldsymbol{y})\} + e^{n\alpha} P_0(\boldsymbol{y}) \exp\{n\theta\hat{H}(\boldsymbol{y})\} \leq e^{n\beta} P_1(\boldsymbol{y})\right\}.$$

With this approach in mind, one can also address directly large deviations constraints:

$$g(\boldsymbol{y}) = P_0(\boldsymbol{y}) \cdot \mathcal{I}\{\boldsymbol{y} :\ \hat{H}(\boldsymbol{y}) \geq R\}$$

and

$$f(\boldsymbol{y}) = P_1(\boldsymbol{y}) \cdot \mathcal{I}\{\boldsymbol{y} :\ \hat{H}(\boldsymbol{y}) \geq R\}.$$

- Universal detection rules can also be devised (see paper).
- For wider classes of sources, replace empirical entropy by LZ complexity.

Summary and Conclusion

- A framework for joint detection and other info processing tasks.
- In general, the main issue is to identify the functions f, g and h.
- Performance analysis – extendable beyond DMS's.