Information Theory

Chapter 4: Information Channels and Capacity

Rudolf Mathar

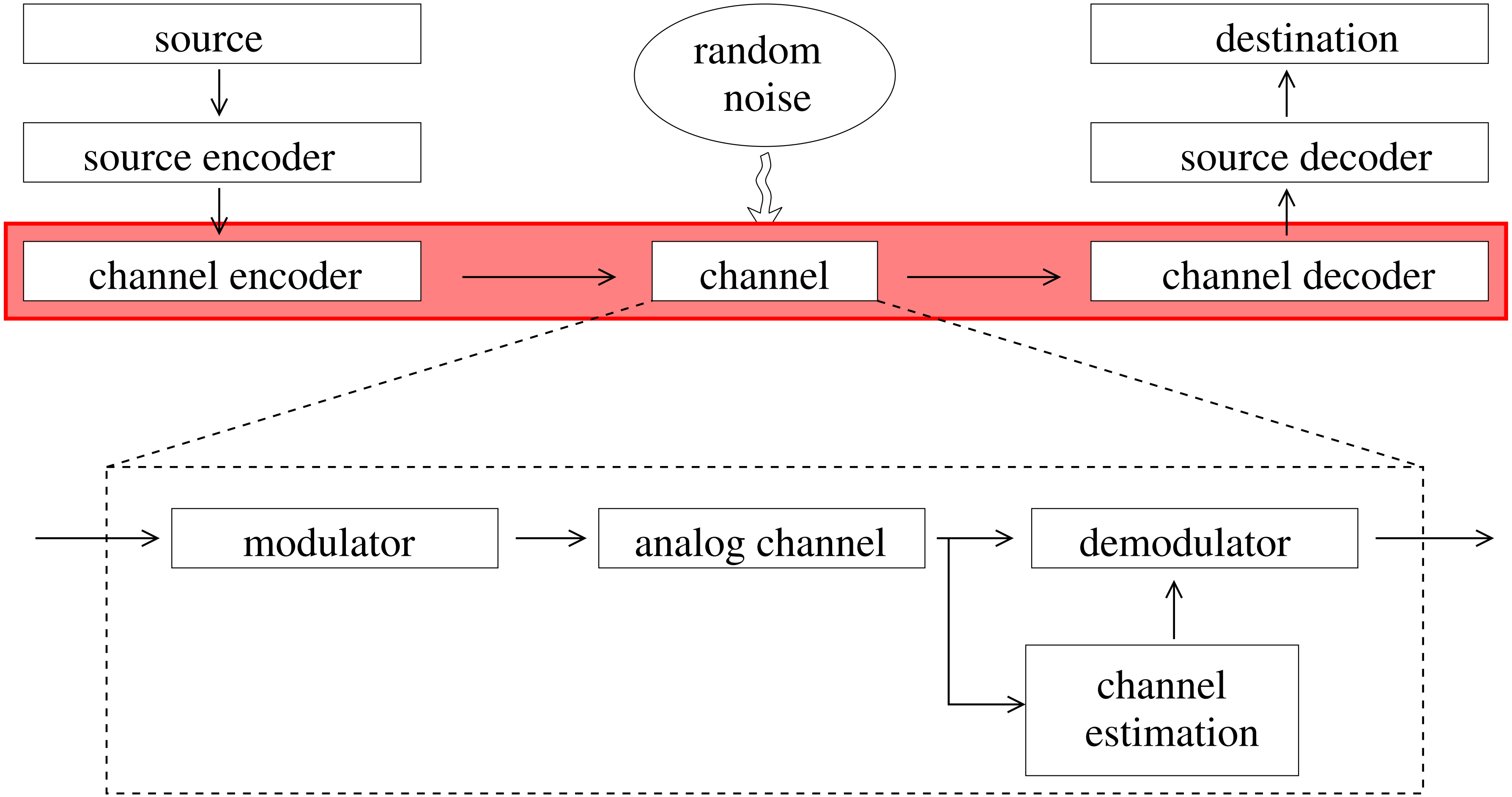
Lehrstuhl für Theoretische Informationstechnik
RWTH Aachen University

WS 2018/19

Outline Chapter 4: Information Channels

Communication Channel

from an information theoretic point of view



Discrete Channel Model

Discrete information channels are described by

- A pair of random variables

$$(X, Y) \text{ with support } \mathcal{X} \times \mathcal{Y},$$

X is the input r.v., $\mathcal{X} = \{x_1, \dots, x_m\}$ the input alphabet.
Y is the outputr.v., $\mathcal{Y} = \{y_1, \dots, y_d\}$ the output alphabet.

- The channel matrix

$$\boldsymbol{W} = \left(w_{ij}\right)_{i=1,\dots,m,\ j=1,\dots,d}$$

with

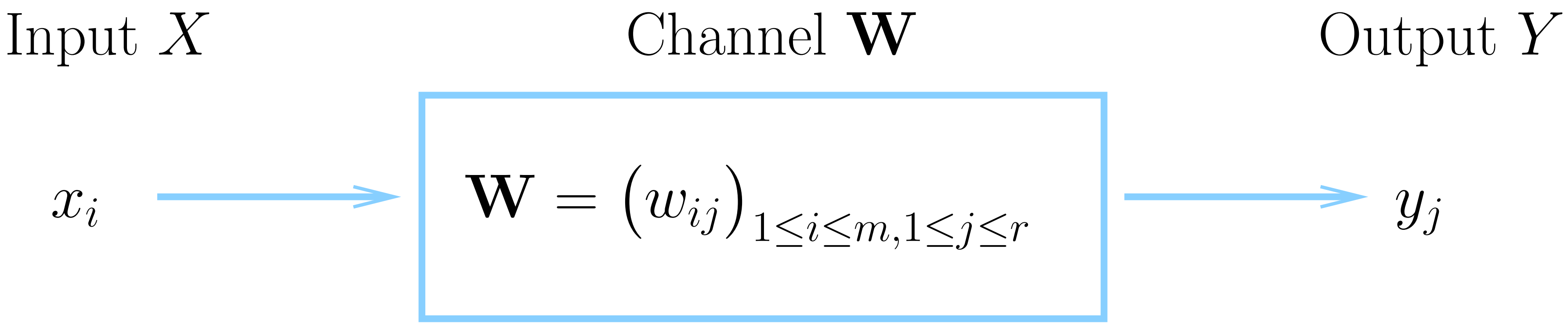
$$w_{ij} = P\left(Y = y_j \mid X = x_i\right),\ \ i = 1, \dots, m,\ j = 1, \dots, d$$

- Input distribution

$$P(X = x_i) = p_i,\ \ i = 1, \dots, m,$$

$\boldsymbol{p} = (p_1, \dots, p_m)$.

Discrete Channel Model



Write $\boldsymbol{W}$ composed of rows $\boldsymbol{w}_1, \dots, \boldsymbol{w}_m$ as $\boldsymbol{W} = \begin{pmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_m \end{pmatrix}$

Lemma 4.1

$$H(Y) = H(\boldsymbol{p}\boldsymbol{W})$$

$$H(Y \mid X = x_i) = H(\boldsymbol{w}_i)$$

$$H(Y \mid X) = \sum_{i=1}^{m} p_i H(\boldsymbol{w}_i)$$

Channel Capacity

Mutual information

$$
\begin{aligned}
I(X;Y) &= H(Y) \quad - H(Y \mid X) \\
&= H(\boldsymbol{p}\boldsymbol{W}) - \sum_{i=1}^{m} p_i H(\boldsymbol{w}_i) \\
&= \sum_{i=1}^{m} p_i D\big(\boldsymbol{w}_i \,\|\, \boldsymbol{p}\boldsymbol{W}\big) = I(\boldsymbol{p};\boldsymbol{W}),
\end{aligned}
$$

D denoting the Kulback-Leibler divergence.

Aim: For a given channel $\boldsymbol{W}$ use the input distribution that maximizes mutual information $I(X;Y)$.

Definition 4.2.

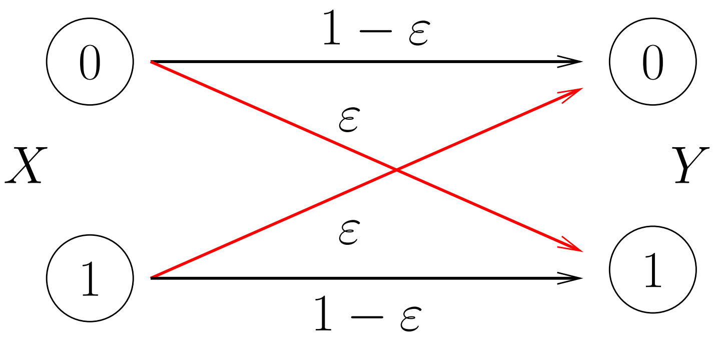
$$
C = \max_{(p_1,\ldots,p_m)} I(X;Y) = \max_{\boldsymbol{p}} I(\boldsymbol{p},\boldsymbol{W})
$$

is called *channel capacity*.

Determining capacity is in general a complicated optimization problem.

Binary Symmetric Channel (BSC)

Example 4.3.



Input distribution $\boldsymbol{p} = (p_0, p_1)$

Channel matrix

$$\boldsymbol{W} = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$$

Mutual Information:

$$I(X;Y) = I(\boldsymbol{p};\boldsymbol{W}) = H(\boldsymbol{p}\boldsymbol{W}) - \sum_{i=1}^{m} p_i H(\boldsymbol{w}_i)$$

$$= \overbrace{H\big(p_0(1-\varepsilon) + p_1\varepsilon, \varepsilon p_0 + (1-\varepsilon)p_1\big)} - \overbrace{H\big(\varepsilon, 1-\varepsilon\big)}$$

The maximum of $I(\boldsymbol{p},\boldsymbol{W})$ over all input distributions (p_0, p_1) is achieved at

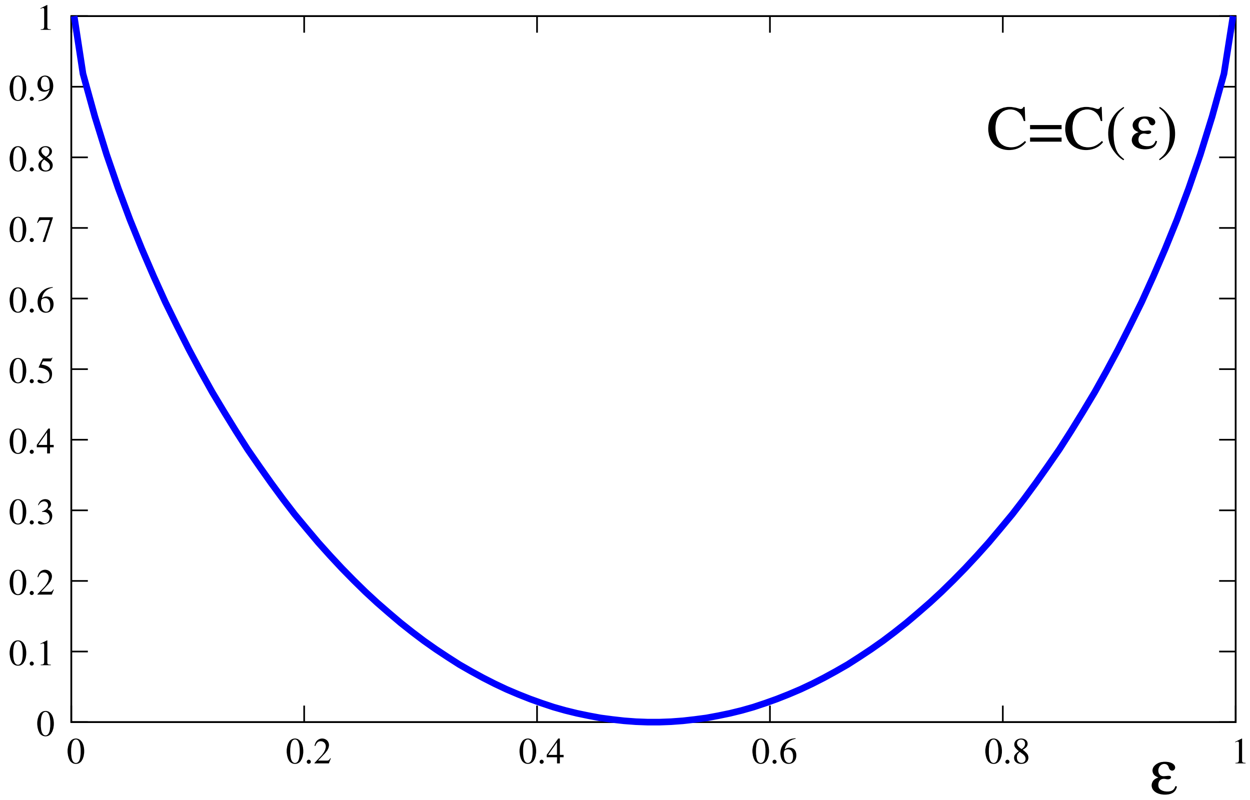
$$(p_0^*, p_1^*) = (0.5, 0.5)$$

Binary Symmetric Channel (BSC)

Hence, $p_0^* = p_1^* = \frac{1}{2}$ is capacity-achieving. It holds

$$C = \max_{(p_0, p_1)} I(X;Y) = 1 + (1-\varepsilon)\log_2(1-\varepsilon) + \varepsilon \log_2 \varepsilon$$

Capacity of the BSC as a function of ε:



Channel Capacity (ctd.)

Given a channel with channel matrix $\boldsymbol{W}$. To compute channel capacity solve

$$C = \max_{\boldsymbol{p}} I(\boldsymbol{p}; \boldsymbol{W}) = \max_{\boldsymbol{p}} \sum_{i=1}^{m} p_i D\big(\boldsymbol{w}_i \,\|\, \boldsymbol{p}\boldsymbol{W}\big)$$

Theorem 4.4.

The capacity of the channel $\boldsymbol{W}$ is attained at $\boldsymbol{p}^* = (p_1^*, \dots, p_m^*)$ if and only if

$$D(\boldsymbol{w}_i \,\|\, \boldsymbol{p}^*\boldsymbol{W}) = \zeta \quad \text{for all} \quad i = 1, \dots, m.$$

for all $i = 1, \dots, m$ with $p_i > 0$.

Moreover,

$$C = I(\boldsymbol{p}^*; \boldsymbol{W}) = \zeta.$$

Channel Capacity (ctd.)

Proof of the Theorem:

Mutual information $I(\boldsymbol{p};\boldsymbol{W})$ is a concave function of $\boldsymbol{p}$. Hence the KKT conditions (cf., e.g., Boyd and Vandenberge 2004) are necessary and sufficient for optimality of some input distribution $\boldsymbol{p}$. Using the above representation some elementary algebra shows that

$$\frac{\partial}{\partial p_i} I(\boldsymbol{p};\boldsymbol{W}) = D(\boldsymbol{w}_i \| \boldsymbol{p}\boldsymbol{W}) - 1.$$

The full set of KKT conditions now reads as

$$\begin{aligned} \textstyle\sum_{j=1}^m p_j &= 1 \\ p_i &\geq 0,\ \ i = 1, \dots, m \\ \lambda_i &\geq 0,\ \ i = 1, \dots, m \\ \lambda_i p_i &= 0,\ \ i = 1, \dots, m \\ D(\boldsymbol{w}_i \| \boldsymbol{p}\boldsymbol{W}) + \lambda_i + \nu &= 0,\ \ i = 1, \dots, m \end{aligned}$$

which shows the assertion.

Channel Capacity (ctd.)

Denote self information by $\rho(q) = -q \log q, \;\; q \geq 0$.

Theorem 4.5. (G. Alirezaei, 2018)
Given a channel with square channel matrix $\boldsymbol{W} = \big(w_{ij}\big)_{i,j=1,\dots,m}$. Assume that $\boldsymbol{W}$ is invertible with inverse

$$\boldsymbol{T} = \big(t_{ij}\big)_{i,j=1,\dots,m}.$$

Then, measured in nats, the capacity is

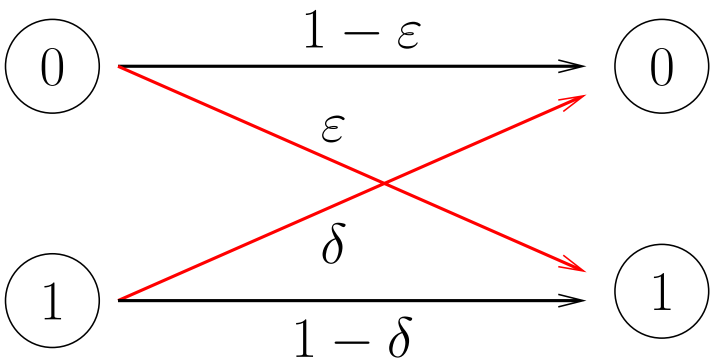
$$C = \ln\Big(\sum_k \exp\{-\sum_{i,j} t_{ki}\, \rho(w_{ij})\}\Big)$$

and the capacity achieving distribution is given by

$$p_\ell^* = e^{-C} \sum_k t_{ks} \exp\{-\sum_{i,j} t_{ki}\, \rho(w_{ij})\} = \frac{\sum_k t_{ks} \exp\{-\sum_{i,j} t_{ki}\, \rho(w_{ij})\}}{\sum_k \exp\{-\sum_{i,j} t_{ki}\, \rho(w_{ij})\}}$$

Binary Asymmetric Channel (BAC)

Example 4.6.



$$\boldsymbol{W} = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \delta & 1-\delta \end{pmatrix}$$

The capacity-achieving distribution is

$$p_0^* = \frac{1}{1+b}, \quad p_1^* = \frac{b}{1+b},$$

with

$$b = \frac{a\varepsilon - (1-\varepsilon)}{\delta - a(1-\delta)} \quad \text{and} \quad a = \exp\left(\frac{h(\delta)-h(\varepsilon)}{1-\varepsilon-\delta}\right),$$

and $h(\varepsilon) = H(\varepsilon, 1-\varepsilon)$, the entropy of $(\varepsilon, 1-\varepsilon)$.

Note that $\varepsilon = \delta$ yields the previous result for the BSC.

Binary Asymmetric Channel (BAC)

Derivation of capacity for the BAC:

By Theorem 4.4 the capacity-achieving input distribution $\boldsymbol{p} = (p_0, p_1)$ satisfies

$$D(\boldsymbol{w}_1 \| \boldsymbol{p}\boldsymbol{W}) = D(\boldsymbol{w}_2 \| \boldsymbol{p}\boldsymbol{W}).$$

This is an equation in the variables p_0, p_1 which jointly with the condition $p_0 + p_1 = 1$ has the solution

$$p_0^* = \frac{1}{1+b}, \quad p_1^* = \frac{b}{1+b}, \tag{1}$$

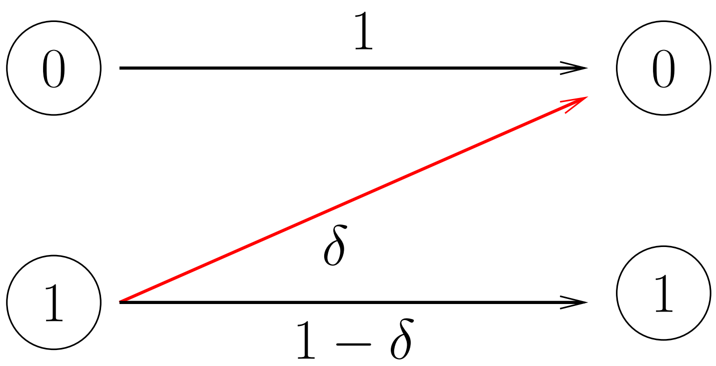
with

$$b = \frac{a\varepsilon - (1-\varepsilon)}{\delta - a(1-\delta)} \quad \text{and} \quad a = \exp\left(\frac{h(\delta) - h(\varepsilon)}{1 - \varepsilon - \delta}\right),$$

and $h(\varepsilon) = H(\varepsilon, 1-\varepsilon)$, the entropy of $(\varepsilon, 1-\varepsilon)$.

Binary Z-Channel (BZC)

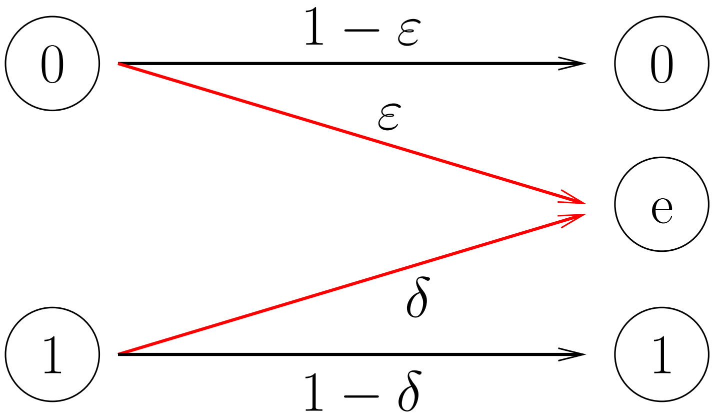
Example 4.7.
The so called Z-channel is a special case of the BAC with $\varepsilon = 0$.



The capacity-achieving distribution is obtained from the BAC by setting $\varepsilon = 0$.

Binary Asymmetric Erasure Channel (BAEC)

Example 4.8.



$$\boldsymbol{W} = \begin{pmatrix} 1-\varepsilon & \varepsilon & 0 \\ 0 & \delta & 1-\delta \end{pmatrix}$$

The capacity-achieving distribution is determined by finding the solution x^* of

$$\varepsilon \log \varepsilon - \delta \log \delta = (1-\delta) \log(\delta + \varepsilon x) - (1-\varepsilon) \log(\varepsilon + \delta / x)$$

and setting

$$\frac{p_0^*}{p_1^*} = x^*, \quad p_0^* + p_1^* = 1.$$

Binary Asymmetric Erasure Channel (BAEC)

Derivation of capacity for the BAEC:

By Theorem 4.4 the capacity-achieving distribution $\boldsymbol{p}^* = (p_0^*, p_1^*)$, $p_0^* + p_1^* = 1$ is given by the solution of

$$\begin{aligned}(1-\varepsilon)\log\frac{1-\varepsilon}{p_0(1-\varepsilon)} + \varepsilon\log\frac{\varepsilon}{p_0\varepsilon + p_1\delta}\\ = \delta\log\frac{\delta}{p_0\varepsilon + p_1\delta} + (1-\delta)\log\frac{1-\delta}{p_0(1-\delta)},\end{aligned} \tag{2}$$

Substituting $x = \frac{p_0}{p_1}$, equation (2) reads equivalently as

$$\varepsilon\log\varepsilon - \delta\log\delta = (1-\delta)\log(\delta + \varepsilon x) - (1-\varepsilon)\log(\varepsilon + \delta/x)$$

By differentiating w.r.t. x it is easy to see that the right hand side is monotonically increasing such that exactly one solution $\boldsymbol{p}^* = (p_1^*, p_2^*)$ exists, which can be numerically computed.