ITCT Lecture 12:
Rate Distortion Function and Optimal Bit-Allocation

Recall:

Information theory

: the problem of source coding- "what information should be sent", and the problem of channel coding- "how should it be sent", can be separated without loss of optimality.

Source coding theorem:

: the entropy H(x) of a source x is the minimum rate at which a source can be encoded without information loss.

Channel coding theorem:

:for error-free transmission over a channel with capacity C, the rate of the source needs to be smaller or equal to C.

The rate distortion function (RFD) R(D)

: the lower bound on the rate necessary to represent a certain source with a given distortion. The RDF can be considered as an extension of the idea of entropy for the case when a certain distortion in the representation of the data is permissible.

If a certain channel capacity C is given, the RDF can be used to find the necessary maximum average distortion D_{ave} so that the condition for error-free transmission $R(D_{ave}) < C$ is achieved. Conversely the RDF can also be used to find the minimum capacity necessary to transmit a certain source with a given average distortion.

Source $\{x_i, p(x_i)\}$

Channel : $Q(y_j \mid x_i)$

The marginal probability of the received message set is

$$T(Y_i) = \sum_j p(x_j) Q(y_i \mid x_j)$$

The average mutual information for a block length N, $I_N(x,y)$ is defined as

$$I_N(x,y) = \sum_i \sum_j p(x_i) Q(y_j \mid x_i) \log \frac{Q(y_j \mid x_i)}{T(y_j)}$$

$$= E_{x,y}\left[\log\left(\frac{1}{T(y_j)} \right) - \log\left(\frac{1}{Q(y_j \mid x_i)} \right) \right],$$

Where Ex,y[·] is the expectation operator w.r.t. x and y.

Mutual information between x,y is the expectation value of the difference between the self information gained by observing a message y_j at the sink and the information gained by observing a message y_j at the sink given that we know that message x_i was sent.

Error-free transmission ➔ knowing x_i implies knowing y_j ➔ y=x and $T(y_i) = P(x_i)$

The information transmitted is equal to the average information of y, which is called the N-th-order entropy of y, i.e.,

$$\underset{Error-free}{I_N(x,y)} = H_N(x) = H_N(y) = \sum_i p(x_i) \log\left(\frac{1}{p(x_i)}\right)$$

Optimal error-free coding requires at least $H_N(x)$ bits.

$$I_N(x, y) = H_N(x) - H_N(x \mid y)$$

Both $I_N(\cdot,\cdot)$ and $H(\cdot)$ are non-negative quantities,

$$I_N(x, y) \le H_N(x)$$

:No information can be added by the encoding stage.

Let d(x,y) be the distortion between x and y. The expected distortion for a given channel Q is:

$$D(Q) = \frac{1}{N} E_{x,y}[d(x, y)]$$

$$= \frac{1}{N} \sum_{i} \sum_{j} p(x_i) Q(y_j \mid x_i) d(x_i, y_j)$$

The N-blocks RDF ($R_N(D_*)$) is the minimum of the average mutual information per symbol, for D(Q) less than some value D_*, that is,

$$R_N(D_*) = \inf_{Q:D(Q)\leq D_*} \frac{1}{N} I_N(x, y)$$

The RDF is obtained by taking the limit as the block length N goes to infinity,

$$R_N(D_*) = \lim_{N \to \infty} R_N(D_*)$$

The RDF is the limit as the block size goes to infinity of the average mutual information per source symbol, subject to the constraint that the average distortion is less than D_* , where the minimization is performed overall encoding scheme as described by the conditional probability assignment $Q(\cdot|\cdot)$

Characteristics of RDF : continuous, convex, differentiable and non-increasing

- The channel capacity is defined by

$$c = \lim_{N \to \infty} \sup_{w} \frac{1}{N} I_N(U, V)$$

Where U is the channel input and V is the channel output and $W(V_j|U_i)$ is the conditional probability assignment which describes the channel mathematically

Source coding theorem for a describe source with a fidelity criterion:

When distortion is permissible, states that if $C > R(D_*)$ then, by using large enough blocks (VQ-based), a source can be sent over a channel with capacity C with an average distortion less than or equal to D_* .

The converse of this theorem is also true:

If $C < R(D_*)$ then it is impossible to send that source over a channel with capacity C having an average distortion less than or equal to D_* .

Remarks : The power of RDF stems from the fact that absolute bounds can be found on the performance of a lossy data compression scheme. Unfortunately, most of the proofs are existence proofs, which means that they prove there exist a bound, but they do not provide a mechanism for constructing a scheme which achieves this bound.

Operational Rate Distortion Theory : (ORDT)

ORDT is based on the fact that every lossy data compression scheme has only a finite set of admissible quantizers.

– There is only a finite number of possible rate distortion pairs for any given source. These pairs form the quantizer function (QF).

Let Q = $\{q_0,\ldots,q_{M-1})$ be the set of all admissible quantizers. Let $R(q_i)$ be the rate for a particular quantizer and a particular source and $D(q_i)$ the corresponding distortion. Then,

$$QF = \{(R(q_i), D(q_i))\}_{i=0}^{M-1}$$

The quantizers which result in the ORDF (Operational Rate Distortion Function) be defined as follows:

$$Q_{ORDF} = \{q : q \in Q, R(q) \geq R(p) \Rightarrow D(q) < D(p), \forall p \in Q\}$$

which is equivalent to the following definitions:

$$Q_{ORDF} = \{q : q \in Q, D(q) \geq D(p) \Rightarrow R(q) < R(p), \forall p \in Q\}$$

: a quantizer can only belong to Q_{ORDF} if there is no other quantizer which results in a lower or equal rate and a lower distortion or equivalently, there is no other quantizer which results in a smaller distortion using the same or a smaller rate.

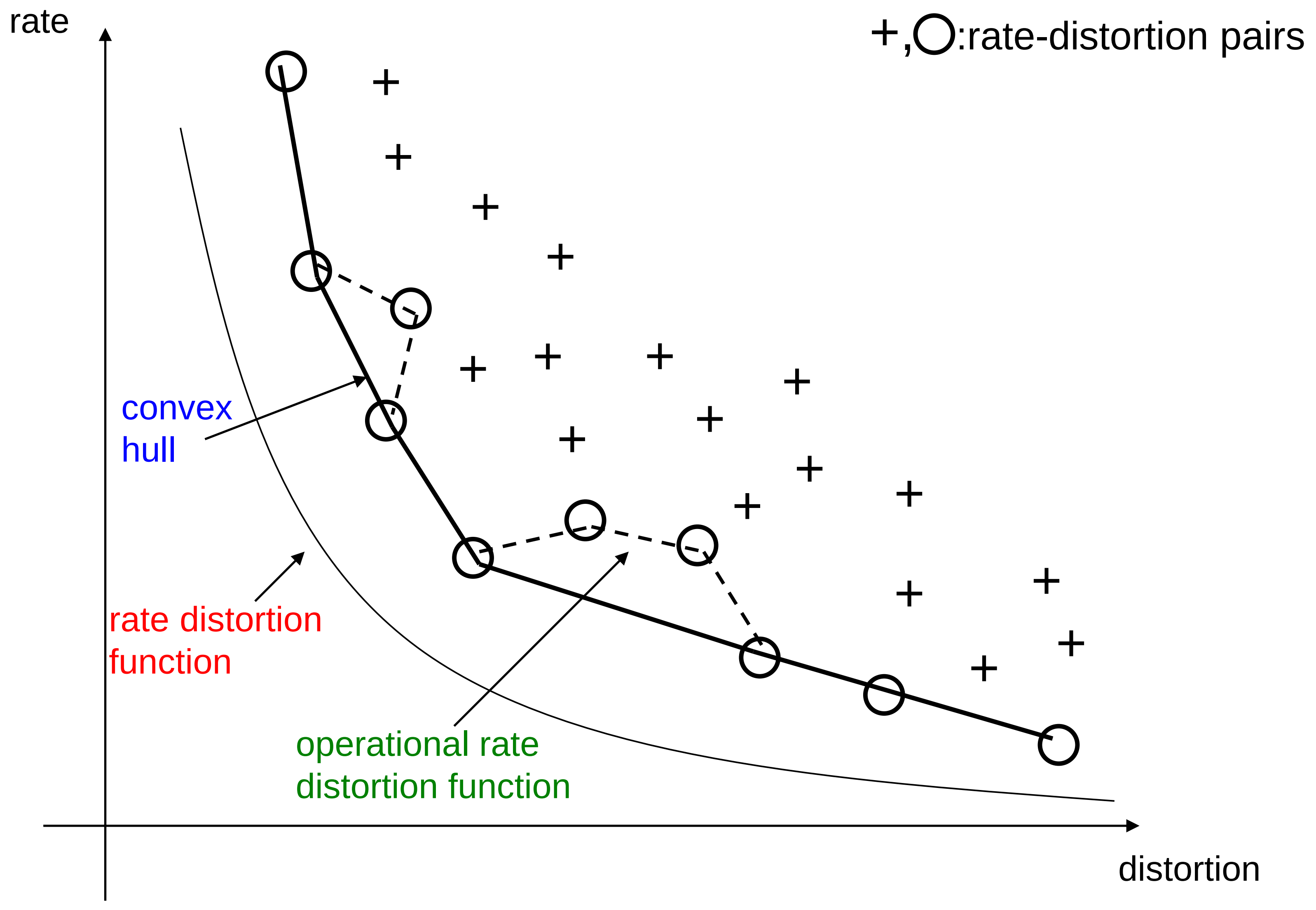
We relabel the quantizers in Q_{ORDF} from zero to $| Q_{ORDF} |-1$ such that $D(q_i) < D(q_{i+1})$, which by definition of Q_{ORDF} implies that $R(q_i) > R(q_{i+1})$

The ORDF is defined as the following ordered set or rate distortion points.

$$ORDF = \left\{\left(R(q_i), D(q_i)\right)\right\}_{i=0}^{|Q_{ORDF}|-1}$$

Remarks:

1. The idea of ORDF is to analyze a given lossy compression scheme and to find its theoretical optimal performance, whereas RDF tries to find the absolute performance bound w.r.t. any scheme.
2. RDF is useful in assessing how close the performance of an actual scheme comes to the theoretical optimal, ORDF can be used to optimize the actual scheme to perform at its best.
   – ORDF is closely related to optimal bit allocation.



rate
+,O:rate-distortion pairs
convex
hull
rate distortion
function
operational rate
distortion function
distortion

Optimal Bit Allocation:

to distribute the available bit budget among different sources of information such that the resulting overall distortion is minimized.

Remarks:

1. Independent v.s. dependent quantizers.
2. MINAVE v.s. MINMAX
3. Real Number v.s. Integer
4. Model-based (RDF) v.s. ORDF.

References:

1. “Bit allocation for dependent quantization with applications to multi-resolution and MPEG video coders,” IEEE Trans. on Image Processing, pp.533-545, Sept. 1994.
2. “Rate-distortion optimal (fast thresholding with complete JPEG/MPEG decoder compatibility, IEEE Trans. on Image Processing, pp. 700-704, Sept. 1994.

A Simple Algorithm:

One starts by giving zero bits to each source and then allocates enough bits to the source with the highest distortion so that its distortion is reduced by the smallest amount possible, using the QF of that source. Then the source with the largest distortion is found and again bits are allocated to that source until its distortion drops. This is repeated until all bits are used up.

Lagrangian Multiplier Method:

Theorem 1:

Let S_B be a finite set and $B \in S_B$ be a member of that set. Let R(B) and D(B) be real valued functions defined over S_B. Then, for any $\lambda \geq 0$ the optimal solution $B^*(\lambda)$ to the unconstrained problem,

$$\min_{B \in S_B} (D(B) + \lambda R(B))$$

is also an optimal solution to the constrained problem,

$$\min_{B \in S_B} (D(B))$$

subject to : $R(B) \leq R(B^*(\lambda))$

Proof. (by contradiction):

Assume that the above theorem is false.

➔ There exists a B $\in S_B$ such that

$D(B) < D(B^*(\lambda))$ and $R(B) \leq R(B^*(\lambda))$.

➔ $D(B) + \lambda \cdot R(B) < D(B^*(\lambda)) + \lambda \cdot R(B^*(\lambda))$,

which is a contradiction since

$B^*(\lambda)$ is the optimal solution to the unconstrained problem.

Remark:

Theorem 1 does not guarantee a solution to the constrained problem. All it says is that to every non-negative λ , there exists a corresponding constrained problem whose solution is identical to that of the unconstrained problem.

That is:

> If $R(B^*(\lambda))$ happens to be equal to an upper bound R_{max}, $B^*(\lambda)$ is the desired solution to the constrained problem. A problem remains to be solved is how to find a
>
> $\lambda \ni R(B^*(\lambda)) = R\text{max}$

Theorem 2: If $R(B^*(\lambda_1)) > R(B^*(\lambda_2))$ then

λ-theorem

$$\lambda_2 \geq -\frac{D(B^*(\lambda_1)) - D(B^*(\lambda_2))}{R(B^*(\lambda_1)) - R(B^*(\lambda_2))} \geq \lambda_1$$

Proof:

By the optimality of $B^*(\lambda_1)$, the following holds,

$$D(B^*(\lambda_1)) + \lambda_1 \cdot R(B(\lambda_1)) \leq D(B^*(\lambda_2)) + \lambda_1 \cdot R(B^*(\lambda_2))$$

Solving for λ_1, and keeping in mind that $R(B^*(\lambda_1) > R(B^*(\lambda_2))$,

$$\Longrightarrow \quad -\frac{D(B^*(\lambda_1)) - D(B^*(\lambda_2))}{R(B^*(\lambda_1)) - R(B^*(\lambda_2))} \geq \lambda_1$$

Similarly, we can prove

$$\lambda_2 \geq -\frac{D(B^*(\lambda_1)) - D(B^*(\lambda_2))}{R(B^*(\lambda_1)) - R(B^*(\lambda_2))}$$

The λ-theorem states that given two optimal solutions, the ratio of the change in optimal distortion to the change in the required rate is bounded between the two multipliers.

➔

If the set of solutions produced by Lagrangian multipliers results in a distortion that is a differentiable function of the rate at some point, then from λ-theorem,

λ at that point is the derivative of the distortion w.r.t. the rate

➔ $$-\frac{dD}{dR} = \lambda$$

By applying this results to the ORDF

(assuming the derivative exists)

assumptions!!

➔ $$\frac{dR}{dD} = -\frac{1}{\lambda}$$

: fast convex search for the optimal λ!

Iterative Search for λ

Theorem 3:

$R(B^*(\lambda))$ is a non-increasing function of the Lagrangian multiplier λ

proof:

Let $\lambda_2 > \lambda_1 > 0$ and assume that $R(B^*(\lambda_2)) > R(B^*(\lambda_1))$. Then from the λ-theorem it follows that $\lambda_1 \geq \lambda_2$, which is a contradiction. That is, $R(B^*(\lambda_2)) < R(B^*(\lambda_1))$

Corollary 1: $D(B^*(\lambda))$ is a non-decreasing function of λ.

Since $R(B^*(\lambda))$ is a non-increasing function of λ, the bisection method can be used to find the optimal λ.

Start:

two initial guesses for λ, λ_l and λ_u which have to be selected $\ni R(B^*(\lambda_l)) \geq R\max \geq R(B^*(\lambda_u))$, where R max is maximum number of bits available i.e., the target bitrate.

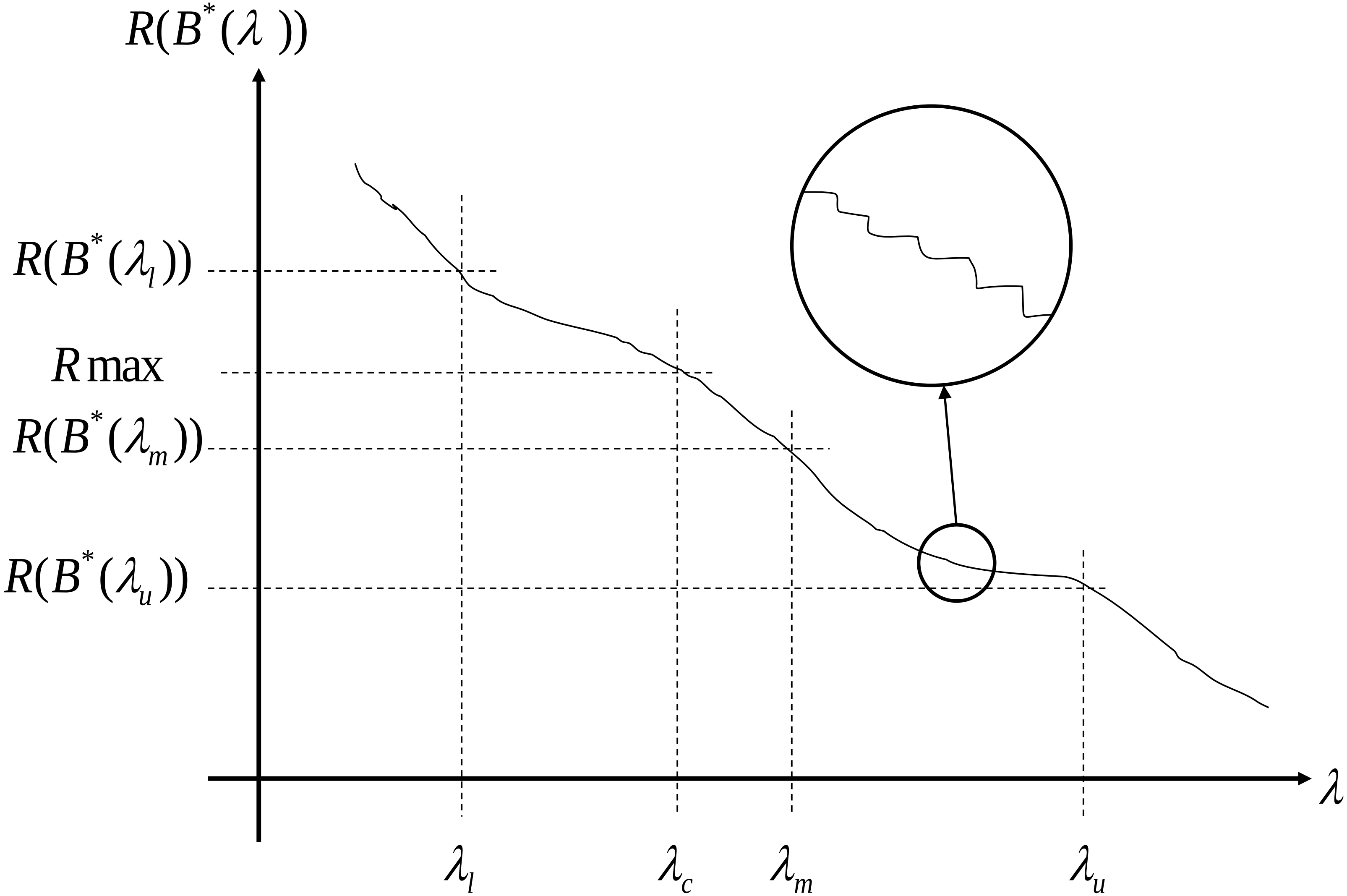
Since $R(B^*(\lambda))$ is non-increasing, the desired λ (if it exists) lies between λ_l and λ_u. By selecting $\lambda_m = \frac{\lambda_l + \lambda_u}{2}$, the initial interval is bisected. Then $R(B^*(\lambda_m))$ is evaluated, and if $R(B^*(\lambda_m)) \geq R\max$ then $\lambda_l = \lambda_m$, otherwise $\lambda_u = \lambda_m$

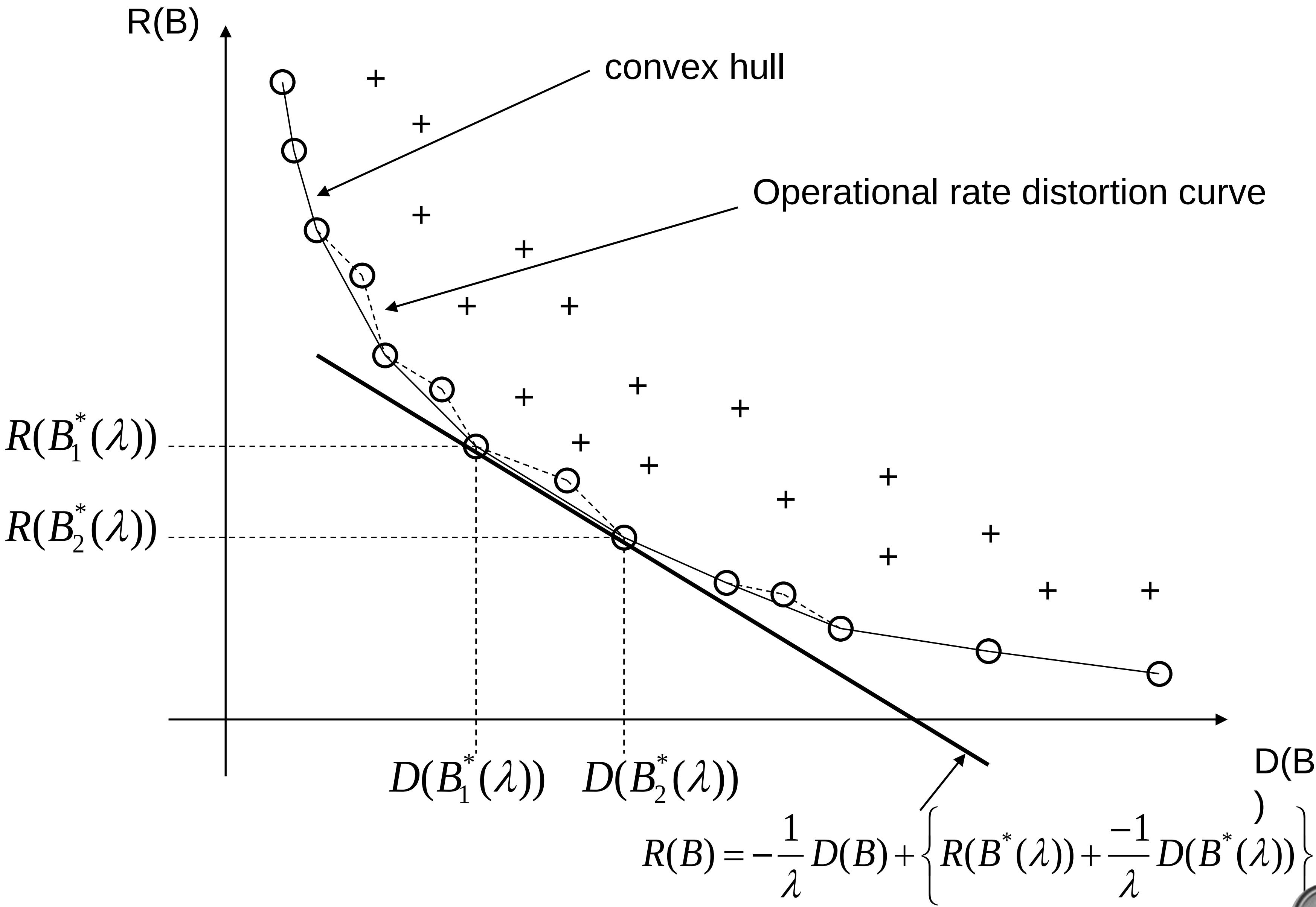
By repeating this procedure, tighter and tighter bounds for the required λ are found.

As mentioned before, there might not be λ which satisfies $R(B^*(\lambda)) = R\max$ exactly. Since R is defined on the finite set S_B. Therefore, the bisection is stopped when a solution is found which satisfies

$$R\max - E \leq R(B^*(\lambda)) \leq R\max + E$$

where $E \geq 0$


$R(B^*(\lambda))$
$R(B^*(\lambda_l))$
$R\max$
$R(B^*(\lambda_m))$
$R(B^*(\lambda_u))$
λ
λ_l
λ_c
λ_m
λ_u



Note that not all circles belong to the convex hull. These points are inaccessible for the Lagrangian multiplier method.

Convex Hull Solutions:

RDF is a non-increasing convex function. ORDF is a set of points in the rate distortion space. When the points of ORDF are connected by straight lines, the resulting curve is not necessarily convex.

In the following, we show that the Lagrangian multiplier method can only be used to find the points of the ORDF which belong to the convex hull of the ORDF.

Assume that $B^*(\lambda)$ is an optimal solution to $\min_{B \in S_B} (D(B) + \lambda R(B))$.

This implies the following,

$$D(B^*(\lambda)) + \lambda \cdot R(B^*(\lambda)) \leq D(B) + \lambda \cdot R(B)$$

Where $B \in S_B$ can be any admissible decision vector.

By rearranging terms, one can obtain

$$R(B) \geq -\frac{1}{\lambda} D(B) + \left\{ R(B^*(\lambda)) + \frac{1}{\lambda} D(B^*(\lambda)) \right\} \quad ...(\Delta)$$

Consider the case when the equal sign in (Δ) holds.

Eqn. (Δ) defines a line in the rate distortion space (plane).

Eqn. (Δ) states that all points (D(B),R(B)) of the QF must lie in the upper right half-plane of the rate distortion plane, i.e., above the line defined by eqn. (Δ), whereas the set of optimal solution $\left\{(D(B_i^*(\lambda)), R(B_i^*(\lambda))\right\}$ is on the line.

Recall that the slope of the line equals $-\frac{1}{\lambda}$, the rate $R(B^*(\lambda))$ is a non-increasing function of λ and the distortion $D(B^*(\lambda))$ is a non-decreasing function of λ.

If we sweep λ from zero to infinity, and connect the points $(D(B^*(\lambda)), R(B^*(\lambda)))$ by straight lines, the convex hull of the ORDF is generated.

A point of the ORDF which is not the convex hull can not be found using a Lagrangian approach since one cannot find a line (and hence a λ) which partitions the plane such that all solutions belong to the resulting upper right half-plane.

For the Lagrangian multiplier method to work, the only requirement is that the optimal solution of the relaxed (unconstrained) problem can be found.

It is important to notice that the role of the rate and the distortion can be interchanged and the presented theory still applies. In other words, the following problem,

$$\min_{B \in S_B} R(B) \quad \text{subject to: } D(B) \leq D\max$$

can also be solved by the Lagrangian multiplier method. (Distortion Rate Function)