Duality for optimization problems with entropy-like objective functions

Radu Ioan Boţ [*] Sorin-Mihai Grad [†] Gert Wanka [‡]

Abstract. We consider a convex optimization problem whose objective function consists of an entropy-like sum of functions $\sum_{i=1}^{k} f_i(x) \ln(f_i(x)/g_i(x))$. We calculate the Lagrange dual of this problem. Weak and strong duality assertions are presented, followed by the derivation of necessary and sufficient optimality conditions. Some entropy optimization problems found in the literature are considered as special cases, their dual problems obtained using other approaches being rediscovered.

Key Words. Entropy optimization, Shannon entropy, Kullback-Leibler entropy, Lagrange dual problem

AMS subject classification (2000). 42A50, 49N15, 54C70

1 Preliminaries

Entropy optimization is a modern and fruitful research area for scientists having various backgrounds: mathematicians, physicists, engineers, even chemists or linguists. Many papers, including two of the present authors ([2] and [3]), and books among which we mention two quite recent ([7] and [10]) deal with entropy optimization, especially with its multitude of applications in various fields such as transport and location problems, pattern and image recognition, text classification, image reconstruction, etc.

The problem we consider here cannot be classified as a pure entropy optimization problem. We may call it a generalization of the usual entropy optimization problems and we argue this statement by the special cases we present in the third part of the present paper.

[*]Faculty of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany, e-mail: radu.bot@mathematik.tu-chemnitz.de.

[†]Faculty of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany, e-mail: grad@mathematik.tu-chemnitz.de. Partially supported by Karl und Ruth Mayer-Stiftung

[‡]Corresponding author. Faculty of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany, e-mail: gert.wanka@mathematik.tu-chemnitz.de

Consider the non-empty convex set $X \subseteq \mathbb{R}^n$, the affine functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., k$, the concave functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., k$, and the convex functions $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, ..., m$. Assume that for $i = 1, ..., k$, $f_i(x) \geq 0$ and $g_i(x) > 0$ when $x \in X$ such that $h(x) \leqq 0$, where $h = (h_1, ..., h_m)^T$ and "$\leqq$" denotes the partial ordering introduced by the corresponding non-negative orthant. Denote further $f = (f_1, ..., f_k)^T$ and $g = (g_1, ..., g_k)^T$. Let us present some other notations we use throughout this paper. All the vectors are column vectors, becoming row-vectors when transposed by an upper index ^{T}. In general, for a vector $u \in \mathbb{R}^p$ we denote its entries by u_j, $j = 1, ..., p$. Moreover, $u > 0$ means actually $u_j > 0\ \forall j = 1, ..., p$. For an optimization problem (U) we denote by $v(U)$ its optimal objective value. For a set A, $\text{int}(A)$ is the interior of the set, while $\text{ri}(A)$ denotes its relative interior. For a matrix $A \in \mathbb{R}^{p \times s}$, A^T denotes the transpose. Moreover, by A_{ij} we denote the entry situated at the intersection of row i and column j in the respective matrix, $i = 1, ..., p$, $j = 1, ..., s$. If there is also another matrix B having the same dimensions as A, $\langle A, B \rangle$ is the inner product between them, being equal to the trace of the product matrix $A^T B$. As expected, for a vector $u \in \mathbb{R}^p$, $\|u\| = \left(\sum_{j=1}^p u_j^2 \right)^{1/2}$ is the Euclidian norm. An important notion that will appear throughout this paper is that of conjugate function, also known as Legendre-Fenchel or Fenchel-Moreau conjugate. For a function $f : \mathbb{R}^p \to \mathbb{R}$, the conjugate function is $f^* : \mathbb{R}^p \to \mathbb{R}$ defined by

$$f^*(u) = \sup_{x \in \mathbb{R}^p} \left\{ u^T x - f(x) \right\}.$$

A direct consequence of this definition is the so-called Fenchel-Young inequality,

$$f^*(u) + f(x) \geq u^T x\ \forall x, u \in \mathbb{R}^p.$$

The convex optimization problem we consider throughout this paper is

$$(P) \qquad \inf_{\substack{x \in X, \\ h(x) \leqq 0}} \left\{ \sum_{i=1}^k f_i(x) \ln \left(\frac{f_i(x)}{g_i(x)} \right) \right\}.$$

As usual in entropy optimization we use further the convention $0 \ln 0 = 0$.

Using a special construction we obtain another problem that is equivalent to (P), whose Lagrange dual problem (D) is easier to determinate. Weak duality between (P) and (D) is certain from the construction, but in order to achieve strong duality we need to introduce a sufficient condition, a so-called constraint qualification. Further we determine some necessary and sufficient optimality conditions regarding the mentioned problems.

The objective function of problem (P) is a Kullback-Leibler-type sum, but instead of probabilities we have as terms functions. To the best of our knowledge this kind of objective function has not been considered yet in the literature.

There are some papers dealing with problems having as objective function expressions like $\int f(t)\ln(f(t)/g(t))$, such as [1], but the results described there do not interfere with ours. Of course the functions involved in the objective function of the problem (P) may take some particular shapes and (P) turns into an entropy optimization problem. The special cases we present later deal with these aspects. For a definite choice of the functions f, g and h and taking $X = \mathbb{R}^n_+$ we obtain the entropy optimization problem with a Kullback-Leibler measure as objective function and convex constraint functions treated in [7]. From (D) we derive a dual problem to this particular one that turns out to be exactly the dual problem obtained via geometric programming in the original paper. When the convex constraint functions h_j, $j = 1, ..., m$, have some more particular properties, i.e. they are linear or quadratic, the dual problems turn into some more specific formulae. As a second special case we took a problem treated by Noll in [11]. After a suitable choice of particular shapes for f, g, h and X we obtain the maximum entropy optimization problem the author used in the applications described in [11], whose objective function is the Shannon entropy of a probability-like vector. The dual problem obtained using Lagrangian duality there arises also when we derive a dual to this problem using (D). A third special case considered here is when the mentioned functions are chosen such that the objective function becomes the so-called Burg entropy minimization problem with linear constraints in [5].

For all the special cases we present the strong duality assertion and necessary and sufficient optimality conditions, derived from the general case.

2 Lagrange Duality for problem (P)

2.1 An equivalent formulation of the problem

The main purpose of the present paper is to study the problem (P) by means of duality. First we need to determine a dual problem to (P), a task rather difficult to accomplish due to the special form of the objective function. Therefore we need to resort to a quite simple construction, similar to the one used by Wanka and Boţ in [14].

Let us introduce the functions $\Phi_i : \mathbb{R}^2 \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$,

$$\Phi_i(s_i, t_i) = \begin{cases} s_i \ln\left(\frac{s_i}{t_i}\right), & s_i \geq 0 \text{ and } t_i > 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

$i = 1, ..., k$, $s = (s_1, ..., s_k)^T$, $t = (t_1, ..., t_k)^T$ and the set

$$\mathcal{A} = \left\{(x, s, t) \in X \times \mathbb{R}^k_+ \times \operatorname{int}(\mathbb{R}^k_+) : h(x) \leqq 0, f(x) = s, t \leqq g(x)\right\}.$$

Now we consider a new optimization problem

$$(P_\Phi) \qquad \inf_{(x,s,t)\in\mathcal{A}} \left\{ \sum_{i=1}^{k} \Phi_i(s_i,t_i) \right\}.$$

For each $i = 1, ..., k$, the function Φ_i is convex, being the extension with positive infinite values to the whole space of a convex function (see [10]). The convexity of the set $\mathcal{A}$ follows from its definition.

Remark: The statements above assure the convexity of the problem (P_Φ).

Even if the problems (P) and (P_Φ) seem related, an accurate connection between their optimal objective values is required. The following assertion states it.

Proposition 1. *The problems* (P) *and* (P_Φ) *are equivalent in the sense that* $v(P) = v(P_\Phi)$.

Proof. Let us take first an element $x \in X$ such that $h(x) \leqq 0$. It is obvious that $(x, f(x), g(x)) \in \mathcal{A}$. Further,

$$\sum_{i=1}^{k} f_i(x) \ln\left(\frac{f_i(x)}{g_i(x)}\right) = \sum_{i=1}^{k} \Phi_i(f_i(x), g_i(x)) \geq v(P_\Phi).$$

As x is chosen arbitrarily in order to fulfill the constraints of the problem (P) we can conclude for the moment that $v(P) \geq v(P_\Phi)$.

Conversely, take a triplet $(x, s, t) \in \mathcal{A}$. This means that we have for each $i = 1, ..., k$, $f_i(x) = s_i$ and $g_i(x) \geq t_i$. Further we have for all $i = 1, ..., k$, $\frac{1}{g_i(x)} \leq \frac{1}{t_i}$, followed by $\frac{f_i(x)}{g_i(x)} \leq \frac{s_i}{t_i}$. Consequently, because ln is a monotonic increasing function, it holds $\ln\left(\frac{f_i(x)}{g_i(x)}\right) \leq \ln\left(\frac{s_i}{t_i}\right)$, $i = 1, ..., k$. Multiplying the terms in both sides by the corresponding $f_i(x) = s_i$, $i = 1, ..., k$, and assembling the resulting relations it follows

$$\sum_{i=1}^{k} \Phi_i(s_i, t_i) \geq \sum_{i=1}^{k} f_i(x) \ln\left(\frac{f_i(x)}{g_i(x)}\right) \geq v(P).$$

As the element (x, s, t) has been taken arbitrarily in $\mathcal{A}$, it yields $v(P_\Phi) \geq v(P)$. Therefore, $v(P) = v(P_\Phi)$. $\square$

Further we determine the Lagrange dual problem of (P_Φ), that is also a dual to problem (P).

2.2 Duality statements concerning problem (P)

As to compute the Lagrange dual problem of (P) is a rather cumbersome task and we have already proved that equivalence stands between problems (P) and (P_Φ), in the following we determine the Lagrange dual problem of the latter. As Lagrange duality is well-known and widely-used, we confine ourselves to proceed with the calculations, without any unnecessary introductions.

So, the Lagrange dual problem to (P_Φ) has the following raw formulation, where q^f, q^g and q^h are the Lagrange multipliers,

$$(D_\Phi) \sup_{\substack{q^f\in\mathbb{R}^k,\\ q^g\in\mathbb{R}^k_+,\\ q^h\in\mathbb{R}^m_+}} \inf_{\substack{x\in X,\\ s\in\mathbb{R}^k_+,\\ t\in\operatorname{int}(\mathbb{R}^k_+)}} \left[\sum_{i=1}^k s_i \ln\left(\frac{s_i}{t_i}\right) + (q^h)^T h(x) + (q^f)^T(f(x)-s) + (q^g)^T(t-g(x))\right].$$

Taking a closer look to the infimum that appears above one may notice that it is separable into a sum of infima in the following way

$$\begin{aligned}
&\inf_{\substack{x\in X,\\ s\in\mathbb{R}^k_+,\\ t\in\operatorname{int}(\mathbb{R}^k_+)}} \left[\sum_{i=1}^k s_i \ln\left(\frac{s_i}{t_i}\right) + (q^h)^T h(x) + (q^f)^T(f(x)-s) + (q^g)^T(t-g(x))\right]\\
=\ & \inf_{x\in X}\left[(q^h)^T h(x) + (q^f)^T f(x) - (q^g)^T g(x)\right]\\
+\ & \sum_{i=1}^k \inf_{\substack{s_i\geq 0,\\ t_i>0}} \left[s_i \ln\left(\frac{s_i}{t_i}\right) - q_i^f s_i + q_i^g t_i\right].
\end{aligned}$$

We can calculate the infima regarding $s_i \geq 0$ and $t_i > 0$ for all $i = 1, ..., k$,

$$\inf_{\substack{s_i\geq 0,\\ t_i>0}} \left[s_i \ln\left(\frac{s_i}{t_i}\right) - q_i^f s_i + q_i^g t_i\right] = \inf_{s_i\geq 0}\left[s_i \ln s_i - q_i^f s_i + \inf_{t_i>0}[q_i^g t_i - s_i \ln t_i]\right].$$

In order to resolve the inner infimum, consider the function $\varphi : (0, +\infty) \to \mathbb{R}$, $\varphi(t) = \alpha t - \beta \ln t$, where $\alpha > 0$ and $\beta \geq 0$. Its minimum is attained at $t = \frac{\beta}{\alpha} > 0$, being $\varphi\left(\frac{\beta}{\alpha}\right) = \beta - \beta\ln\beta + \beta\ln\alpha$. Applying this result to the infima concerning t_i in the expressions above for $i = 1, ..., k$, there follows

$$\inf_{t_i>0}[q_i^g t_i - s_i \ln t_i] = \begin{cases} s_i - s_i\ln s_i + s_i \ln q_i^g, & \text{if } q_i^g > 0,\\ 0, & \text{if } q_i^g = 0 \text{ and } s_i = 0,\\ -\infty, & \text{if } q_i^g = 0 \text{ and } s_i > 0.\end{cases}$$

Further we have to calculate for each $i = 1, ..., k$ the infimum above with respect to $s_i \geq 0$ after replacing the infimum concerning t_i with its value. In case $q_i^g > 0$

we have

$$\begin{aligned}\inf_{s_i\geq 0}\left[s_i\ln s_i - q_i^f s_i + s_i - s_i\ln s_i + s_i\ln q_i^g\right] &= \inf_{s_i\geq 0}\left[s_i(1-q_i^f+\ln q_i^g)\right]\\ &= \begin{cases}0, & \text{if } 1-q_i^f+\ln q_i^g\geq 0,\\ -\infty, & \text{otherwise.}\end{cases}\end{aligned}$$

When $q_i^g = 0$ the infimum concerning s_i is equal to $-\infty$.

One may conclude for each $i = 1, ..., k$, the following

$$\inf_{\substack{s_i\geq 0,\\ t_i>0}}\left[s_i\ln\left(\frac{s_i}{t_i}\right) - q_i^f s_i + q_i^g t_i\right] = \begin{cases}0, & \text{if } 1-q_i^f+\ln q_i^g\geq 0 \text{ and } q_i^g>0,\\ -\infty, & \text{otherwise.}\end{cases} \tag{1}$$

The negative infinite values are not relevant to the dual problem we work on since after determining the inner infima one has to calculate the supremum of the obtained values, so we must consider further the cases where the infima with respect to $s_i \geq 0$ and $t_i > 0$ are 0, i.e. the following constraints have to be fulfilled $1 - q_i^f + \ln q_i^g \geq 0$ and $q_i^g > 0$, $i = 1, ..., k$. The former additional constraints are equivalent to $q_i^g \geq e^{q_i^f - 1}$ $\forall i = 1, ..., k$. Let us write now the final form of the dual problem to (P_Φ), after noticing that as $e^{q_i^f-1} > 0$ the constraints $q^g \in \mathbb{R}_+^k$ and $q_i^g > 0$, $i = 1, ..., k$, become redundant and may be ignored,

$$(D) \qquad \sup_{\substack{q^f\in\mathbb{R}^k, q^h\in\mathbb{R}_+^m,\\ q_i^g\geq e^{q_i^f-1}, i=1,...,k}} \inf_{x\in X}\left[(q^h)^T h(x) + (q^f)^T f(x) - (q^g)^T g(x)\right].$$

Although (D) has been obtained via Lagrangian duality from (P_Φ) we refer to it further as the dual problem to (P) since (P) and (P_Φ) are equivalent. Next we present the duality assertions regarding (P) and (D) beginning with the weak duality statement.

Theorem 2. *There is weak duality between (P) and (D), i.e. $v(P) \geq v(D)$.*

Proof. Theorem 5.1 in [6] yields $v(P_\Phi) \geq v(D)$, so Proposition 1 leads to the conclusion. □

Weak duality always holds, but we cannot assert the same about strong duality. Alongside the initial convexity assumptions for X and h_i, $i = 1, ..., m$, the concavity of g_i, $i = 1, ..., k$, and the affinity of the functions f_i, $i = 1, ..., k$, an additional constraint qualification is sufficient in order to achieve strong duality. The one we use here is inspired from the regularity condition proposed in [6],

$$(CQ) \qquad \exists x' \in \mathrm{ri}(X) : \begin{cases} f(x') > 0,\\ h_j(x') \leq 0, & \text{if } j \in L,\\ h_j(x') < 0, & \text{if } j \in N,\end{cases}$$

where we have divided the set $\{1, ..., m\}$ into two disjunctive sets as follows

$$L = \Big\{j \in \{1, ..., m\} : h_j \text{ is the restriction to } X \text{ of an affine function}\Big\}$$

and $N = \{1, ..., m\} \backslash L$. The strong duality statement arises naturally.

Theorem 3. *If the constraint qualification* (CQ) *is fulfilled then there is strong duality between problems* (P) *and* (D)*, i.e.* (D) *has an optimal solution and* $v(P) = v(P_\Phi) = v(D)$.

Proof. Since $\Phi_i(s_i, t_i) \geq 0 \ \forall (x, s, t) \in \mathcal{A}$ (the most important properties of the Kullback-Leibler entropy measure are presented and proved in [10]) it follows

$$v(P_\Phi) \geq 0. \tag{2}$$

The constraint qualification (CQ) being fulfilled, there is a triplet $(x', s', t') \in \mathrm{ri}(X) \times \mathrm{int}(R^k_+) \times \mathrm{int}(R^k_+)$ such that

$$\begin{cases} h_j(x') \leq 0, & j \in L, \\ h_j(x') < 0, & j \in N, \\ f(x') = s', & \\ t'_i < g_i(x'), & i = 1, ..., k. \end{cases} \tag{3}$$

For instance take $s' = f(x')$ and $t' = \frac{1}{2} g(x')$.

The results (2) and (3) allow us to apply Theorem 5.7 in [6], so strong duality between (P_Φ) and (D) is certain, i.e. (D) has an optimal solution and $v(P_\Phi) = v(D)$. Proposition 1 yields $v(P) = v(D)$. $\square$

Another step forward is to present some necessary and sufficient optimality conditions regarding the pair of dual problems we treat.

Theorem 4. *(a) Let the constraint qualification* (CQ) *be fulfilled and assume that the primal problem* (P) *has an optimal solution* $\bar{x}$*. Then the dual problem* (D) *has an optimal solution, too, let it be* $(\bar{q}^f, \bar{q}^g, \bar{q}^h)$*, and the following optimality conditions are true,*

(i) $f_i(\bar{x}) \ln \left(\frac{f_i(\bar{x})}{g_i(\bar{x})} \right) = \bar{q}^f_i f_i(\bar{x}) - \bar{q}^g_i g_i(\bar{x}), \ i = 1, ..., k,$

(ii) $\inf\limits_{x \in X} \Big[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x) \Big] = (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x}),$

(iii) $\bar{q}^h_j h_j(\bar{x}) = 0, \ j = 1, ..., m.$

(b) If $\bar{x}$ *is a feasible point to* (P) *and* $(\bar{q}^f, \bar{q}^g, \bar{q}^h)$ *is feasible to* (D) *fulfilling the optimality conditions* $(i) - (iii)$*, then there is strong duality between* (P) *and* (D).

Moreover, $\bar{x}$ is an optimal solution to the primal problem and $(\bar{q}^f, \bar{q}^g, \bar{q}^h)$ an optimal solution to the dual.

Proof. (a) Under weaker assumptions than here Theorem 3 yields strong duality between (P) and (D). Therefore the existence of an optimal solution $(\bar{q}^f, \bar{q}^g, \bar{q}^h)$ to the dual problem is guaranteed. Moreover, $v(P) = v(D)$ and because (P) has an optimal solution its optimal objective value is attained at $\bar{x}$ and we have

$$\sum_{i=1}^{k} f_i(\bar{x}) \ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right) = \inf_{x\in X}\left[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x)\right] \tag{4}$$

Earlier we have proved the validity of (1). Using it we can determine the conjugate function of Φ_i, $i = 1, ..., k$, at $\left(\bar{q}_i^f, -\bar{q}_i^g\right)$ as follows

$$\begin{aligned}
\Phi_i^*(\bar{q}_i^f, -\bar{q}_i^g) &= \sup_{(s_i,t_i)\in\mathbb{R}^2}\left\{(\bar{q}_i^f, -\bar{q}_i^g)^T(s_i,t_i) - \Phi_i(s_i,t_i)\right\}\\
&= \sum_{i=1}^{k} \sup_{s_i\geq 0, t_i>0}\left\{\bar{q}_i^f s_i - \bar{q}_i^g t_i - s_i \ln\left(\frac{s_i}{t_i}\right)\right\}\\
&= -\sum_{i=1}^{k} \inf_{\substack{s_i\geq 0,\\ t_i>0}}\left[s_i \ln\left(\frac{s_i}{t_i}\right) - \bar{q}_i^f s_i + \bar{q}_i^g t_i\right]\\
&= \begin{cases} 0, & \text{if } 1 - \bar{q}_i^f + \ln \bar{q}_i^g \geq 0 \text{ and } \bar{q}_i^g > 0,\\ +\infty, & \text{otherwise.}\end{cases}
\end{aligned}$$

As $\bar{q}^f$ and $\bar{q}^g$ are feasible to (D) we have $\Phi_i^*(\bar{q}_i^f, -\bar{q}_i^g) = 0$ $\forall i = 1, ..., k$. Let us apply Young's inequality for $\Phi_i(f_i(\bar{x}), g_i(\bar{x}))$ and $\Phi_i^*(\bar{q}_i^f, -\bar{q}_i^g)$, when $i = 1, ..., k$. We have

$$\Phi_i(f_i(\bar{x}), g_i(\bar{x})) + \Phi_i^*(\bar{q}_i^f, -\bar{q}_i^g) \geq \bar{q}_i^f f_i(\bar{x}) - \bar{q}_i^g g_i(\bar{x}),\ i = 1, ..., k.$$

Summing these relations up one gets

$$\sum_{i=1}^{k} f_i(\bar{x}) \ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right) \geq (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x}). \tag{5}$$

On the other hand it is obvious that

$$\inf_{x\in X}\left[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x)\right] \leq (\bar{q}^h)^T h(\bar{x}) + (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x}). \tag{6}$$

Relations $(4) - (6)$ yield

$$\begin{aligned}
0 &= \sum_{i=1}^{k} f_i(\bar{x}) \ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right) - \inf_{x\in X}\left[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x)\right]\\
&\geq (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x}) - \left[(\bar{q}^h)^T h(\bar{x}) + (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x})\right]\\
&= -(\bar{q}^h)^T h(\bar{x}) \geq 0.
\end{aligned}$$

The last inequality holds due to the fact that $\bar{x}$ is feasible to (P) and $\bar{q}^h$ to (D). Thus, of course all of these inequalities must be fulfilled as equalities. Therefore we immediately have (iii) and

$$\sum_{i=1}^{k}\left[\quad f_i(\bar{x})\ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right)-\left((\bar{q}_i^f)^T f_i(\bar{x})-(\bar{q}_i^g)^T g_i(\bar{x})\right)\right]+\Big[(\bar{q}^h)^T h(\bar{x})+(\bar{q}^f)^T f(\bar{x})$$
$$-(\bar{q}^g)^T g(\bar{x})\Big]-\inf_{x\in X}\Big[(\bar{q}^h)^T h(x)+(\bar{q}^f)^T f(x)-(\bar{q}^g)^T g(x)\Big]=0.$$

This yields the fulfillment of the above Young's inequality for Φ_i and Φ_i^* as equality, that is nothing but (i). With (iii) then also (ii) is clear.

(b) The conclusion arises obviously following the proof above backwards. $\square$

3 Applications

This section is dedicated to some interesting special cases of the problem treated so far. The first of them is the convex-constrained minimum cross-entropy problem, then follows a norm-constrained maximum entropy problem and as a third special case we present a so-called linearly constrained Burg entropy optimization problem. The cross-entropy problem has been treated so far by means of geometric duality in [7] and some other papers, for the second Noll determined the Lagrange dual problem in [11], while the Burg entropy problem we present comes from [5]. The fact that the dual problems we obtain in the first two special cases are actually the ones determined in the original papers shows that the problem we treated is a generalization of the classical entropy optimization problems.

3.1 The Kullback-Leibler entropy as objective function

The book [7] is a must for anyone interested in entropy optimization. Among many other interesting statements and applications, the authors consider the cross-entropy minimization problem with convex constraint functions

$$(P_K)\qquad \inf_{\substack{x\in\mathbb{R}^n_+,\ \sum\limits_{i=1}^n x_i=1,\\ l_j(A_jx)+b_j^Tx+c_j\leq 0,\\ j=1,...,r}}\left\{\sum_{i=1}^n x_i\ln\left(\frac{x_i}{q_i}\right)\right\},$$

where A_j are $k_j\times n$ matrices with full row-rank, $b_j\in\mathbb{R}^n$, $j=1,...,r$, $c=(c_1,...,c_r)^T\in\mathbb{R}^r$, $l_j:\mathbb{R}^{k_j}\to\mathbb{R}$, $j=1,...,r$, are convex functions and there is also the probability distribution $q=(q_1,...,q^n)^T\in\text{int}(\mathbb{R}^n_+)$, with $\sum\limits_{i=1}^n q_i=1$. We omit the additional assumptions of differentiability and co-finiteness for the functions

$l_j,\ j = 1, ..., r$ from the mentioned paper.

After dealing with the problem (P_K) we particularize its constraints like in [7], first to become linear, then to obtain a quadratically-constrained cross-entropy optimization problem. We determine their dual problems and present the corresponding strong duality assertions and optimality conditions.

The problem (P_K) is a special case of our problem (P) when the elements involved are taken as follows

$$\begin{cases} X = \mathbb{R}^n_+, k = n, m = r + 2, \\ f_i(x) = x_i \ \forall x \in \mathbb{R}^n, i = 1, ..., n, \\ g_i(x) = q_i \ \forall x \in \mathbb{R}^n, i = 1, ..., n, \\ h_j(x) = l_j(A_j x) + b_j^T x + c_j \ \forall x \in \mathbb{R}^n, j = 1, ..., m-2, \\ h_{m-1}(x) = \sum\limits_{i=1}^n x_i - 1 \ \forall x \in \mathbb{R}^n, \\ h_m(x) = 1 - \sum\limits_{i=1}^n x_i \ \forall x \in \mathbb{R}^n. \end{cases}$$

We want to determine the dual problem to (P_K) which is to be obtained from (D) by replacing the terms involved with the above-mentioned expressions. Let us proceed

$$\begin{aligned} &(q^f)^T f(x) = (q^f)^T x, \quad (q^g)^T g(x) = (q^g)^T q, \\ &(q^h)^T h(x) = \sum_{j=1}^r q_j^h \big(l_j(A_j x) + b_j^T x + c_j\big) + q_{m-1}^h \Big(\sum_{i=1}^n x_i - 1\Big) + q_m^h \Big(1 - \sum_{i=1}^n x_i\Big). \end{aligned}$$

Denoting $w = q_{m-1}^h - q_m^h$, the dual problem to (P_K) is

$$\begin{aligned} (D_K) \sup_{\substack{q^f \in \mathbb{R}^n, w \in \mathbb{R}, q^h \in \mathbb{R}^r_+ \\ q_i^g \geq e^{q_i^f - 1}, i=1,...,n}} \inf_{x \in \mathbb{R}^n_+} \Bigg[& (q^f)^T x - (q^g)^T q + \sum_{j=1}^r q_j^h l_j(A_j x) \\ + & \left(\sum_{j=1}^r q_j^h b_j\right)^T x + (q^h)^T c + w \left(\sum_{i=1}^n x_i - 1\right)\Bigg], \end{aligned}$$

where $\Big(\sum\limits_{j=1}^r q_j^h b_j\Big)_i$, $i = 1, ..., n$, is the i-th entry of the vector $\sum\limits_{j=1}^r q_j^h b_j$. We can rearrange the terms and the dual problem becomes

$$\begin{aligned} (D_K) \sup_{\substack{q^f \in \mathbb{R}^n, w \in \mathbb{R}, q^h \in \mathbb{R}^r_+ \\ q_i^g \geq e^{q_i^f - 1}, i=1,...,n}} \Bigg\{ & \inf_{x \in \mathbb{R}^n_+} \left[\sum_{i=1}^n x_i \Big(q_i^f + \Big(\sum_{j=1}^r q_j^h b_j\Big)_i + w \Big) + \sum_{j=1}^r q_j^h l_j(A_j x) \right] \\ + & (q^h)^T c - w - (q^g)^T q \Bigg\}. \end{aligned}$$

Let us calculate separately the infimum over $x \in \mathbb{R}^n_+$. In order to do this we introduce the linear operators $\tilde{A}_j : \mathbb{R}^n \to \mathbb{R}^{k_j}$ defined by $\tilde{A}_j(x) = A_j x$, $j = 1, ..., m$. So these infima become

$$\inf_{x \in \mathbb{R}^n_+} \left[\sum_{i=1}^n x_i \Big(q_i^f + \Big(\sum_{j=1}^r q_j^h b_j \Big)_i + w \Big) + \sum_{j=1}^r \big((q_j^h l_j) \circ \tilde{A}_j \big)(x) \right]. \tag{7}$$

By Proposition 5.7 in [6] the expression in (7) is equal to

$$\sup_{\gamma \in \mathbb{R}^n_+} \left\{ \inf_{x \in \mathbb{R}^n} \left[\sum_{i=1}^n x_i \Big(q_i^f + \Big(\sum_{j=1}^r q_j^h b_j \Big)_i + w - \gamma_i \Big) + \sum_{j=1}^r \big((q_j^h l_j) \circ \tilde{A}_j \big)(x) \right] \right\}, \tag{8}$$

further equivalent to

$$\sup_{\gamma \in \mathbb{R}^n_+} \left\{ - \sup_{x \in \mathbb{R}^n} \left[\sum_{i=1}^n x_i \Big(\gamma_i - q_i^f - \Big(\sum_{j=1}^r q_j^h b_j \Big)_i - w \Big) - \sum_{j=1}^r \big((q_j^h l_j) \circ \tilde{A}_j \big)(x) \right] \right\}.$$

The inner supremum may be written as a conjugate function, so the term above becomes

$$\sup_{\gamma \in \mathbb{R}^n_+} \left\{ - \Big(\sum_{j=1}^r \big((q_j^h l_j) \circ \tilde{A}_j \big) \Big)^* (\gamma - u) \right\},$$

where we have denoted by u the vector $\Big(q_i^f + \Big(\sum\limits_{j=1}^r q_j^h b_j \Big)_i + w \Big)_{i=1,...,n}$. We can now apply Theorem 16.4 in [12] since the effective domains of the functions $\big(q_j^h l_j\big) \circ \tilde{A}_j$, $j = 1, ..., r$, coincide, being equal to $\mathbb{R}^n$. We have

$$\Big(\sum_{j=1}^r \big((q_j^h l_j) \circ \tilde{A}_j \big) \Big)^* (\gamma - u) = \inf_{\substack{a_j \in \mathbb{R}^n, j=1,...,r, \\ \sum\limits_{j=1}^r a_j = \gamma - u}} \left[\sum_{j=1}^r \big((q_j^h l_j) \circ \tilde{A}_j \big)^* (a_j) \right]. \tag{9}$$

The relation (7) is now equivalent to

$$\sup_{\gamma \in \mathbb{R}^n_+} \left\{ - \inf_{\substack{a_j \in \mathbb{R}^n, j=1,...,r, \\ \sum\limits_{j=1}^r a_j = \gamma - u}} \left[\sum_{j=1}^r \big((q_j^h l_j) \circ \tilde{A}_j \big)^* (a_j) \right] \right\}$$

and furthermore to

$$\sup_{\substack{a_j \in \mathbb{R}^n, j=1,...,r, \\ \gamma \in \mathbb{R}^n_+, \sum\limits_{j=1}^r a_j = \gamma - u}} \left\{ - \sum_{j=1}^r \big((q_j^h l_j) \circ \tilde{A}_j \big)^* (a_j) \right\}.$$

As for any $j = 1, ..., r$, the image set of the operator $\tilde{A}_j$ is included into $\mathbb{R}^{k_j}$ that is the domain of the function $q_j^h l_j$ defined by $\big(q_j^h l_j\big)(x) = q_j^h l_j(x)$, we may apply Theorem 16.3 in [12] and the last expression becomes equivalent to

$$\sup_{\substack{a_j \in \mathbb{R}^n, j=1,...,r, \\ \gamma \in \mathbb{R}^n_+, \sum\limits_{j=1}^r a_j = \gamma - u}} \left\{ -\sum_{j=1}^r \inf_{\substack{\lambda_j \in \mathbb{R}^{k_j}, \\ \tilde{A}_j^* \lambda_j = a_j}} \left[\big(q_j^h l_j\big)^*(\lambda_j) \right] \right\}, \tag{10}$$

where $\tilde{A}_j^*$ is the adjoint of the operator $\tilde{A}_j$, $j = 1, ..., r$. Turning the inner infima into suprema and drawing all the variables under the leading supremum (10) is equivalent, after applying the definition of the adjoint of a linear operator, to

$$\sup_{\substack{\gamma \in \mathbb{R}^n_+, a_j \in \mathbb{R}^n, \lambda_j \in \mathbb{R}^{k_j}, j=1,...,r, \\ \sum\limits_{j=1}^r a_j = \gamma - u, A_j^T \lambda_j = a_j}} \left\{ -\sum_{j=1}^r \big(q_j^h l_j\big)^*(\lambda_j) \right\}.$$

One may remark that the variables γ and a_j, $j = 1, ..., r$, are superfluous, so the expression is further simplifiable to

$$\sup_{\substack{\lambda_j \in \mathbb{R}^{k_j}, j=1,...,r, \\ \left(\sum\limits_{j=1}^r A_j^T \lambda_j + u\right) \in \mathbb{R}^n_+}} \left\{ -\sum_{j=1}^r \big(q_j^h l_j\big)^*(\lambda_j) \right\}.$$

Let us resume the calculations concerning the dual problem using the partial results obtained above. The dual problem to (P_K) becomes

$$(D_K) \qquad \sup_{\substack{q^f \in \mathbb{R}^n, q_i^g \geq e^{q_i^f - 1}, i=1,...,n, \\ w \in \mathbb{R}, q^h \in \mathbb{R}^r_+, \lambda_j \in \mathbb{R}^{k_j}, j=1,...,r, \\ q_i^f + \left(\sum\limits_{j=1}^r q_j^h b_j\right)_i + w + \left(\sum\limits_{j=1}^r A_j^T \lambda_j\right)_i \geq 0, \\ i=1,...,n,}} \left\{ (q^h)^T c - w - (q^g)^T q - \sum_{j=1}^r \big(q_j^h l_j\big)^*(\lambda_j) \right\},$$

rewritable as

$$(D_K) \qquad \sup_{\substack{q^f \in \mathbb{R}^n, w \in \mathbb{R}, q^h \in \mathbb{R}^r_+, \lambda_j \in \mathbb{R}^{k_j}, j=1,...,r, \\ q_i^f + \left(\sum\limits_{j=1}^r q_j^h b_j\right)_i + w + \left(\sum\limits_{j=1}^r A_j^T \lambda_j\right)_i \geq 0, \\ i=1,...,n,}} \left\{ (q^h)^T c - w - \sum_{j=1}^r \big(q_j^h l_j\big)^*(\lambda_j) + \sum_{i=1}^n \sup_{q_i^g \geq e^{q_i^f - 1}} \left\{ -q_i^g q_i \right\} \right\}.$$

It is obvious that $\sup\big\{ -q_i^g q_i : q_i^g \geq e^{q_i^f - 1}\big\} = -q_i e^{q_i^f - 1}$, $i = 1, ..., n$, so the dual problem turns into

$$(D_K) \quad \sup_{\substack{q^f \in \mathbb{R}^n, w \in \mathbb{R}, q^h \in \mathbb{R}^r_+, \lambda_j \in \mathbb{R}^{k_j}, j=1,...,r, \\ q_i^f + \left(\sum\limits_{j=1}^r q_j^h b_j\right)_i + w + \left(\sum\limits_{j=1}^r A_j^T \lambda_j\right)_i \geq 0, \\ i=1,...,n,}} \left\{ (q^h)^T c - w - \sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j) - \sum_{i=1}^n q_i e^{q_i^f - 1} \right\}.$$

The suprema after q_i^f, $i = 1, ..., n$, are easily computable since the constraints are linear inequalities and the objective functions are monotonic decreasing, i.e.

$$\sup\left\{ -e^{q_i^f - 1} : q_i^f + \left(\sum_{j=1}^r \left(q_j^h b_j + A_j^T \lambda_j\right)\right)_i + w \geq 0 \right\} = -e^{-w - \left(\sum\limits_{j=1}^r \left(q_j^h b_j + A_j^T \lambda_j\right)\right)_i - 1}.$$

Back to the dual problem, it becomes

$$(D_K) \quad \sup_{\substack{w \in \mathbb{R}, q^h \in \mathbb{R}^r_+, \\ \lambda_j \in \mathbb{R}^{k_j}, j=1,...,r}} \left\{ (q^h)^T c - \sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j) - w - \sum_{i=1}^n q_i e^{-w - \left(\sum\limits_{j=1}^r \left(q_j^h b_j + A_j^T \lambda_j\right)\right)_i - 1} \right\}.$$

The next variable we want to renounce is w. In order to do this let us consider the function $\eta : \mathbb{R} \to \mathbb{R}$, $\eta(w) = -w - Be^{-w-1}$, $B > 0$. Its derivative is $\eta'(w) = Be^{-w-1} - 1$, $w \in \mathbb{R}$, a monotonic decreasing function that annulates at $w = \ln B - 1$. So η attains its maximal value at $w = \ln B - 1$, that is $\eta(\ln B - 1) = -\ln B$. Applying these considerations to our dual problem for $B = \sum\limits_{i=1}^n q_i e^{-\left(\sum\limits_{j=1}^r \left(q_j^h b_j + A_j^T \lambda_j\right)\right)_i}$ we get rid of variable $w \in \mathbb{R}$ and the simplified version of the dual problem is

$$(D_K) \quad \sup_{\substack{q^h \in \mathbb{R}^r_+, \lambda_j \in \mathbb{R}^{k_j}, \\ j=1,...,r}} \left\{ (q^h)^T c - \sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j) - \ln\left(\sum_{i=1}^n q_i e^{-\left(\sum\limits_{j=1}^r \left(q_j^h b_j + A_j^T \lambda_j\right)\right)_i} \right) \right\},$$

that turns out, after redenoting the variables, to be the dual problem obtained in [7] via geometric duality.

As weak duality between (P_K) and (D_K) is certain, we focus on the strong duality. In order to achieve it we particularize the constraint qualification (CQ) as follows

$$(CQ_K) \qquad \exists x' > 0 : \begin{cases} \sum\limits_{i=1}^n x_i' = 1, \\ l_j(A_j x') + b_j^T x' + c_j \leq 0, & \text{if } j \in L_K, \\ l_j(A_j x') + b_j^T x' + c_j < 0, & \text{if } j \in N_K, \end{cases}$$

where the sets L_K and N_K are defined analogously to L and K, i.e.

$$L_K = \big\{j \in \{1,...,r\} : l_j \text{ is an affine function}\big\} \text{ and } N_K = \{1,...,r\}\backslash L_K.$$

We are ready now to enunciate the strong duality assertion.

Theorem 5. *If the constraint qualification (CQ_K) is fulfilled, then there is strong duality between problems (P_K) and (D_K), i.e. (D_K) has an optimal solution and $v(P_K) = v(D_K)$.*

Proof. From the general case we have strong duality between (P_K) and the first formulation of the dual problem in this section. The equality $v(P_K) = v(D_K)$ has been preserved after all the steps we performed in order to simplify the formulation of the dual, but there could be a problem regarding the existence of the solution to the dual problem. Fortunately, the results applied to obtain (8), (9) and (10) mention also the existence of a solution to the resulting problems, respectively, so this property is preserved up to the final formulation of the dual problem. □

Furthermore we give also some necessary and sufficient optimality conditions in the following statement. They were obtained in the same way as in Theorem 4, so we have decided to omit the proof, avoiding an unnecessary lengthening of the paper.

Theorem 6. *(a) Let the constraint qualification (CQ_K) be fulfilled and assume that the primal problem (P_K) has an optimal solution $\bar{x}$. Then the dual problem (D_K) has an optimal solution, too, let it be $(\bar{q}^h, \bar{\lambda}_1, ..., \bar{\lambda}_r)$, and the following optimality conditions are true,*

$$\text{(i)}\ \sum_{i=1}^{n} \bar{x}_i \ln\left(\frac{\bar{x}_i}{q_i}\right) + \ln\left(\sum_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r}\left(\bar{q}_j^h b_j + A_j^T \bar{\lambda}_j\right)\right)_i}\right) = -\left(\sum_{j=1}^{r}\left(\bar{q}_j^h b_j + A_j^T \bar{\lambda}_j\right)\right)^T \bar{x},$$

$$\text{(ii)}\ \left(\bar{q}_j^h l_j\right)^*(\bar{\lambda}_j) + \left(\bar{q}_j^h l_j\right)\left(A_j \bar{x}\right) = \bar{\lambda}_j^T A_j \bar{x},\ j = 1, ..., r,$$

$$\text{(iii)}\ \bar{q}_j^h\left(l_j(A_j\bar{x}) + b_j^T\bar{x} + c_j\right) = 0,\ j = 1, ..., r.$$

(b) If $\bar{x}$ is a feasible point to (P_K) and $(\bar{q}^h, \bar{\lambda}_1, ..., \bar{\lambda}_r)$ is feasible to (D_K) fulfilling the optimality conditions $(i) - (iii)$, then there is strong duality between (P_K) and (D_K). Moreover, $\bar{x}$ is an optimal solution to the primal problem and $(\bar{q}^h, \bar{\lambda}_1, ..., \bar{\lambda}_r)$ an optimal solution to the dual.

The problem (P_K) may be particularized even more, in order to fit a wide range of applications. We present further two special cases obtained from (P_K) by assigning some particular values to the constraint functions, as indicated also in [7].

3.1.1 Special case 1: Kullback-Leibler entropy objective function and linear constraints

Taking $l_j(y_j) = 0$, $y_j \in \mathbb{R}^{k_j}$, $j = 1, ..., r$, we have for the conjugates involved in the dual problem

$$\left(q_j^h l_j\right)^*(\lambda_j) = \sup_{y_j \in \mathbb{R}^{k_j}} \left\{\lambda_j^T y_j - 0\right\} = \begin{cases} 0, & \lambda_j = 0, \\ +\infty, & \text{otherwise,} \end{cases} \quad j = 1, ..., r.$$

Performing the necessary substitutions, we get the following pair of dual problems

$$(P_L) \qquad \inf_{\substack{x \in \mathbb{R}^n_+, \sum\limits_{i=1}^{n} x_i = 1, \\ b_j^T x + c_j \leq 0, j = 1, ..., r}} \left\{ \sum_{i=1}^{n} x_i \ln\left(\frac{x_i}{q_i}\right) \right\}$$

and

$$(D_L) \qquad \sup_{\substack{q^h \in \mathbb{R}^r_+, \\ j = 1, ..., r}} \left\{ (q^h)^T c - \ln\left\{ \sum_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r} q_j^h b_j\right)_i} \right\} \right\}.$$

In [7] there is treated a similar problem to (P_L), but instead of inequality constraints Fang et al. use equality constraints. The dual problem they obtain is also similar to (D_L), the only difference consisting of the feasible set, $\mathbb{R}^r_+$ to (D_L), respectively $\mathbb{R}^r$ in [7]. Let us mention further that an interesting application of the optimization problem with Kullback-Leibler entropy objective function and linear constraints can be found in [8]. In order to achieve strong duality the sufficient constraint qualification is

$$(CQ_L) \qquad \exists x' > 0 : \begin{cases} \sum\limits_{i=1}^{n} x_i' = 1, \\ b_j^T x' + c_j \leq 0, \quad j = 1, ..., r. \end{cases}$$

Theorem 7. *If the constraint qualification (CQ_L) is valid, then there is strong duality between problems (P_L) and (D_L), i.e. (D_L) has an optimal solution and $v(P_L) = v(D_L)$.*

As this assertion is a special case of Theorem 5 we omit its proof. The optimality conditions arise also easily from Theorem 6.

Theorem 8. *(a) Assume that the primal problem (P_L) has an optimal solution $\bar{x}$ and that the constraint qualification (CQ_L) is fulfilled. Then the dual problem (D_L) has an optimal solution, too, let it be $\bar{q}^h$ and the following optimality conditions are true,* *(i)* $\sum\limits_{i=1}^{n} \bar{x}_i \ln\left(\frac{\bar{x}_i}{q_i}\right) + \ln\left(\sum\limits_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r}\left(\bar{q}_j^h b_j\right)\right)_i}\right) = -\left(\sum\limits_{j=1}^{r} \bar{q}_j^h b_j\right)^T \bar{x},$

(ii) $\bar{q}_j^h\left(b_j^T \bar{x} + c_j\right) = 0,\ j = 1, ..., r.$

(b) If $\bar{x}$ is a feasible point to (P_L) and $\bar{q}^h$ a feasible point to (D_L) fulfilling the optimality conditions (i) and (ii), then there is strong duality between (P_L) and (D_L). Moreover, $\bar{x}$ is an optimal solution to the primal problem and $\bar{q}^h$ one to the dual.

3.1.2 Special case 2: Kullback-Leibler entropy objective function and quadratic constraints

This time assign the following expressions to some elements in (P_K), $l_j(y_j) = \frac{1}{2}y_j^T y_j$, $y_j \in \mathbb{R}^{k_j}$, $j = 1, ..., r$. We have (cf. [12])

$$\left(q_j^h l_j\right)^*(\lambda_j) = \begin{cases} \frac{\|\lambda_j\|^2}{2q_j^h}, & \text{if } q_j^h \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The pair of dual problems is in this case are

$$(P_Q) \qquad \inf_{\substack{x \in \mathbb{R}^n_+, \sum\limits_{i=1}^{n} x_i = 1, \\ \frac{1}{2}x^T A_j^T A_j x + b_j^T x + c_j \leq 0, \\ j=1,...,r}} \left\{ \sum_{i=1}^{n} x_i \ln\left(\frac{x_i}{q_i}\right) \right\}$$

and

$$(D_Q) \quad \sup_{\substack{q^h \in \mathbb{R}^r_+, \lambda_j \in \mathbb{R}^{k_j}, \\ j=1,...,r}} \left\{ (q^h)^T c - \ln\left(\sum_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r}\left(q_j^h b_j + A_j^T \lambda_j\right)\right)_i}\right) - \frac{1}{2}\sum_{j=1}^{r} \frac{\|\lambda_j\|^2}{q_j^h} \right\},$$

exactly the one in [7].

The following constraint qualification is sufficient in order to assure strong duality

$$(CQ_Q) \qquad \exists x' > 0 : \begin{cases} \sum\limits_{i=1}^{n} x'_i = 1, \\ \frac{1}{2}x'^T A_j^T A_j x' + b_j^T x' + c_j < 0, \quad j = 1, ..., r. \end{cases}$$

Theorem 9. *If the constraint qualification (CQ_Q) is fulfilled, then there is strong duality between problems (P_Q) and (D_Q), i.e. (D_Q) has an optimal solution and $v(P_Q) = v(D_Q)$.*

Furthermore, we give without proof also some necessary and sufficient optimality conditions in the following statement.

Theorem 10. *(a) Let the constraint qualification (CQ_Q) be fulfilled and assume that the primal problem (P_Q) has an optimal solution $\bar{x}$. Then the dual problem (D_Q) has an optimal solution, too, let it be $(\bar{q}^h, \bar{\lambda}_1, ..., \bar{\lambda}_r)$ and the following optimality conditions are true,*

(i) $\sum\limits_{i=1}^{n} \bar{x}_i \ln\left(\frac{\bar{x}_i}{q_i}\right) + \ln\Big(\sum\limits_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r}\left(\bar{q}_j^h b_j + A_j^T \bar{\lambda}_j\right)\right)_i}\Big) = -\Big(\sum\limits_{j=1}^{r}\left(\bar{q}_j^h b_j + A_j^T \bar{\lambda}_j\right)\Big)^T \bar{x}$,

(ii) $\frac{1}{2}\bar{q}_j^h \bar{x}^T A_j^T A_j \bar{x} + \frac{\|\bar{\lambda}_j\|^2}{2\bar{q}_j^h} = \bar{\lambda}_j^T A_j \bar{x}$, $j = 1, ..., r$,

(iii) $\bar{q}_j^h\left(\bar{x}^T A_j^T A_j \bar{x} + b_j^T \bar{x} + c_j\right) = 0$, $j = 1, ..., r$.

(b) If $\bar{x}$ is a feasible point to (P_K) and $(\bar{q}^h, \bar{\lambda}_1, ..., \bar{\lambda}_r)$ is feasible to (D_K) fulfilling the optimality conditions $(i) - (iii)$, then there is strong duality between (P_K) and (D_K). Moreover, $\bar{x}$ is an optimal solution to the primal problem, while $(\bar{q}^h, \bar{\lambda}_1, ..., \bar{\lambda}_r)$ turns out to be an optimal solution to the dual.

3.2 The Shannon entropy as objective function

Noll [11] presents an interesting application of the maximum entropy optimization in image reconstruction considering the following problem

$$(P_S) \quad \inf_{\substack{x_{ij}\geq 0, i=1,...,n, j=1,...m,\\ \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_{ij}=T,\ \|Ax-y\|\leq\varepsilon}} \left\{\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}\ln x_{ij}\right\},$$

where $x \in \mathbb{R}^{n\times m}$ with entries x_{ij}, $i = 1, ..., n$, $j = 1, ..., m$, $A \in \mathbb{R}^{n\times n}$, $y \in \mathbb{R}^{n\times m}$, $\varepsilon > 0$ and $T = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} y_{ij} > 0$. It is easy to notice that the objective function in this problem is the well-known Shannon entropy measure with variables x_{ij}, $i = 1, ..., n$, $j = 1, ...m$, so (P_S) is actually equivalent to the following classical maximum entropy optimization problem

$$(P'_S) \qquad - \sup_{\substack{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_{ij}=T,\ \|Ax-y\|\leq\varepsilon,\\ x_{ij}\geq 0, i=1,...,n, j=1,...m}} \left\{-\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}\ln x_{ij}\right\}.$$

However, (P_S) is viewable as a special case of problem (P) by assigning to the sets and functions involved there the following terms

$$\begin{cases} X = \mathbb{R}_+^{n\times m}, x = (x_{ij})_{\substack{i=1,...,n,\\ j=1,...,m}} \in \mathbb{R}_+^{n\times m}, \\ f_{ij}(x) = x_{ij} \ \forall x \in \mathbb{R}^{n\times m}, i = 1,...,n, j = 1,...m, \\ g_{ij}(x) = 1 \ \forall x \in \mathbb{R}^{n\times m}, i = 1,...,n, j = 1,...m, \\ h_1(x) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_{ij} - T \ \forall x \in \mathbb{R}^{n\times m}, \\ h_2(x) = T - \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_{ij} \ \forall x \in \mathbb{R}^{n\times m}, \\ h_3(x) = \|Ax - y\| - \varepsilon \ \forall x \in \mathbb{R}^{n\times m}. \end{cases}$$

Remark: Some may object that (P_S) is not a pure special case of (P) because the variable x is not an n-dimensional vector as in (P), but a $n \times m$ matrix. As matrices can be viewed also as vectors, in this case the variable becomes an $n \times m$-dimensional vector, so we may apply the results obtained for (P) also to (P_S).

To obtain the dual problem to (P_S) from (D) we calculate the following expressions, where the Lagrange multipliers are now $q^f \in \mathbb{R}^{n\times m}$, $q^g \in \mathbb{R}_+^{n\times m}$ and $q^h \in \mathbb{R}_+^3$,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}^f f_{ij}(x) = \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}^f x_{ij}, \quad \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}^g g_{ij}(x) = \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}^g,$$
$$\sum_{j=1}^{3} q_j^h h_j(x) = \big(q_1^h - q_2^h\big)\left(\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} - T\right) + q_3^h\big(\|Ax - y\| - \varepsilon\big).$$

The multipliers q_1^h and q_2^h appear only together, so we may replace both of them, i.e. their difference, with a new variable $w = q_1^h - q_2^h \in \mathbb{R}$. The dual problem to (P_S) becomes (cf. (D) in section 2.2)

$$(D_S) \quad \sup_{\substack{q^f \in \mathbb{R}^{n\times m},\\ q_3^h \geq 0, w \in \mathbb{R},\\ q_{ij}^g \geq e^{q_{ij}^f} - 1,\\ i=1,...,n, j=1,...,m}} \inf_{x=(x_{ij})_{ij} \in \mathbb{R}_+^{n\times m}} \left[\sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}^f x_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}^g + w\Big(\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} - T\Big) + q_3^h\big(\|Ax - y\| - \varepsilon\big)\right],$$

rewritable as

$$(D_S) \quad \sup_{\substack{q^f \in \mathbb{R}^{n\times m},\\ q_3^h \geq 0, w \in \mathbb{R},\\ q_{ij}^g \geq e^{q_{ij}^f} - 1,\\ i=1,...,n, j=1,...,m}} \left\{ -wT - \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}^g - q_3^h \varepsilon + \inf_{x \in \mathbb{R}_+^{n\times m}} \left[\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}\big(q_{ij}^f + w\big) + q_3^h \|Ax - y\|\right]\right\}.$$

We determine now the infimum concerning $x \in \mathbb{R}^{n\times m}_+$ as in the previous section. By Theorem 16.3 in [12] it turns out to be equal to

$$-\sup_{\substack{q^f_{ij}+w+\left(A^T\Lambda\right)_{ij}\geq 0,\\ i=1,\dots,n,j=1,\dots,m,\Lambda\in\mathbb{R}^{n\times m}}} \left\{\left(q^h_3\|\cdot-y\|\right)^*(\Lambda)\right\}.$$

For the conjugate of the norm function we have (cf. [13])

$$\left(q^h_3\|\cdot-y\|\right)^*(\Lambda)=\begin{cases}-\langle y,\Lambda\rangle, & \text{if } \|\Lambda\|\leq q^h_3,\\ -\infty, & \text{otherwise.}\end{cases}$$

As negative infinite values are not relevant to our problem since there is a leading supremum to be determined, the dual problem becomes

$$(D_S)\quad \sup_{\substack{q^f\in\mathbb{R}^{n\times m},q^h_3\geq 0,w\in\mathbb{R},\Lambda\in\mathbb{R}^{n\times m},\\ \|\Lambda\|\leq q^h_3,q^f_{ij}+w+\left(A^T\Lambda\right)_{ij}\geq 0,\\ q^g_{ij}\geq e^{q^f_{ij}}-1,i=1,\dots,n,j=1,\dots,m}} \left\{-wT-\sum_{i=1}^n\sum_{j=1}^m q^g_{ij}-q^h_3\varepsilon-\langle y,\Lambda\rangle\right\},$$

equivalent to

$$(D_S)\quad \sup_{\substack{q^f\in\mathbb{R}^{n\times m},\\ w\in\mathbb{R},\Lambda\in\mathbb{R}^{n\times m},\\ q^f_{ij}+w+\left(A^T\Lambda\right)_{ij}\geq 0,\\ i=1,\dots,n,j=1,\dots,m}} \left\{-wT-\langle y,\Lambda\rangle+\sum_{i=1}^n\sum_{j=1}^m \sup_{q^g_{ij}\geq e^{q^f_{ij}-1}}\left\{-q^g_{ij}\right\}+\varepsilon\sup_{q^h_3\geq\|\Lambda\|}\left\{-q^h_3\right\}\right\}.$$

The suprema from inside are trivially determinable, so we obtain for the dual problem the following expression

$$(D_S)\quad \sup_{\substack{q^f\in\mathbb{R}^{n\times m},w\in\mathbb{R},\Lambda\in\mathbb{R}^{n\times m},\\ q^f_{ij}+w+\left(A^T\Lambda\right)_{ij}\geq 0,i=1,\dots,n,j=1,\dots,m}} \left\{-wT-\langle y,\Lambda\rangle-\sum_{i=1}^n\sum_{j=1}^m e^{q^f_{ij}-1}-\varepsilon\|\Lambda\|\right\},$$

further equivalent to

$$(D_S)\quad \sup_{\substack{w\in\mathbb{R},\\ \Lambda\in\mathbb{R}^{n\times m}}} \left\{-wT-\langle y,\Lambda\rangle-\varepsilon\|\Lambda\|+\sum_{i=1}^n\sum_{j=1}^m \sup_{q^f_{ij}+w+\left(A^T\Lambda\right)_{ij}\geq 0}\left\{-e^{q^f_{ij}-1}\right\}\right\}.$$

For the inner suprema we have for all $i=1,\dots,n$, and $j=1,\dots,m$,

$$\sup\left\{-e^{q^f_{ij}-1}:q^f_{ij}+w+\left(A^T\Lambda\right)_{ij}\geq 0\right\}=-e^{-w-(A^T\Lambda)_{ij}-1},$$

so the dual problem is simplifiable even to

$$(D_S) \sup_{\substack{w \in \mathbb{R}, \\ \Lambda \in \mathbb{R}^{n \times m}}} \left\{ - wT - \langle y, \Lambda \rangle - \varepsilon \|\Lambda\| - e^{-w-1} \sum_{i=1}^{n} \sum_{j=1}^{m} e^{-(A^T \Lambda)_{ij}} \right\},$$

that is exactly the dual problem obtained via Lagrangian duality in [11].

Moreover, one may notice that also the variable $w \in \mathbb{R}$ could be eradicated. Using the results regarding the maximal value of the function η introduced before, we have

$$\sup_{w \in \mathbb{R}} \left\{ - wT - e^{-w-1} \sum_{i=1}^{n} \sum_{j=1}^{m} e^{-(A^T \Lambda)_{ij}} \right\} = T \left(\ln T - \ln \left(\sum_{i=1}^{n} \sum_{j=1}^{m} e^{-(A^T \Lambda)_{ij}} \right) \right).$$

The last version of the dual problem we reach is

$$(D_S) \sup_{\Lambda \in \mathbb{R}^{n \times m}} \left\{ T \left(\ln T - \ln \left(\sum_{i=1}^{n} \sum_{j=1}^{m} e^{-(A^T \Lambda)_{ij}} \right) \right) - \langle y, \Lambda \rangle - \varepsilon \|\Lambda\| \right\}.$$

As weak duality is certain, we skip it and focus on the strong duality. In order to achieve it the following constraint qualification is sufficient

$$(CQ_S) \qquad \exists x' > 0 : \begin{cases} \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} x'_{ij} = T, \\ \|Ax' - y\| < \varepsilon. \end{cases}$$

The strong duality assertion comes immediately and the necessary and sufficient optimality conditions follow thereafter. Even if the original paper does not contain such statements, we omit the both proofs because they are modifications of the former proofs in the present paper.

Theorem 11. *Assume the constraint qualification (CQ_S) fulfilled. Then strong duality between (P_S) and (D_S) is valid, i.e. (D_S) has an optimal solution and $v(P_S) = v(D_S)$.*

Theorem 12. *(a) Assume the constraint qualification (CQ_S) fulfilled and let $\bar{x}$ be an optimal solution to (P_S). Then the dual problem (D_S) has an optimal solution $\bar{\Lambda}$ and the following optimality conditions hold*

(i) $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} \bar{x}_{ij} \ln \bar{x}_{ij} + T \left(\ln \left(\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} e^{-(A^T \bar{\Lambda})_{ij}} \right) - \ln T \right) = \langle A^T \bar{\Lambda}, \bar{x} \rangle,$

(ii) $\|A\bar{x} - y\| = \varepsilon,$

(iii) $\langle \bar{\Lambda}, A\bar{x} - y \rangle = \|\bar{\Lambda}\| \|A\bar{x} - y\|.$

(b) If $\bar{x}$ is a feasible point to (P_S) and $\bar{\Lambda}$ one to (D_S) satisfying the optimality conditions $(i)-(iii)$, then they are actually optimal solutions to the corresponding problems that enjoy moreover strong duality.

3.3 The Burg entropy as objective function

A third widely-used entropy measure is the one introduced by J.P. Burg. Although there are some others in the literature, we confine ourselves to the most used three, as they have proved to be the most important from the viewpoint of applications. The Burg entropy problem we have chosen as the third application comes from Censor and Lent's paper [5] having Burg entropy as objective function and linear equality constraints,

$$(P_B) \qquad \sup_{\substack{x \in \operatorname{int}(\mathbb{R}^n_+), \\ Ax=b}} \left\{ \sum_{i=1}^n \ln x_i \right\},$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Some other problems with Burg entropy objective function and linear constraints that slightly differ from the one we treat are available, for example, in [4] and [9]. None of these authors gives explicitly a dual to the Burg entropy problem they consider.

The problem (P_B) may be equivalently rewritten as a minimization problem as follows

$$(P_B') \qquad - \inf_{\substack{x \in \operatorname{int}(\mathbb{R}^n_+), \\ Ax=b}} \left\{ - \sum_{i=1}^n \ln x_i \right\}.$$

Denoting (P_B'') the problem (P_B') after eluding the leading minus, it may be trapped as a special case of (P) by taking

$$\begin{cases} X = \operatorname{int}(\mathbb{R}^n_+), k = n, \\ f_i(x) = 1 \ \forall x \in \mathbb{R}^n, i = 1, ..., n, \\ g_i(x) = x_i \ \forall x \in \mathbb{R}^n, i = 1, ..., n, \\ h_1(x) = Ax - b \ \forall x \in \mathbb{R}^n, \\ h_2(x) = b - Ax \ \forall x \in \mathbb{R}^n. \end{cases}$$

To calculate the dual problem to (P_B'') let us replace the values above in (D). We get

$$(D_B) \quad \sup_{\substack{q^f \in \mathbb{R}^n, q_1^h, q_2^h \in \mathbb{R}^m_+, \\ q_i^g \geq e^{q_i^f - 1}, i=1,...,n}} \inf_{x>0} \left[\left(q_1^h - q_2^h\right)^T \left(Ax - b\right) + \sum_{i=1}^n \left(q_i^f - q_i^g x_i\right) \right].$$

Again, we introduce a new variable $w = q_1^h - q_2^h \in \mathbb{R}^m$ to replace the difference of the two positive ones that appear only together. After rearranging the terms the dual becomes

$$(D_B) \quad \sup_{\substack{q^f \in \mathbb{R}^n, w \in \mathbb{R}^m, \\ q_i^g \geq e^{q_i^f - 1}, i=1,...,n}} \left\{ \sum_{i=1}^n q_i^f - w^T b + \sum_{i=1}^n \inf_{x_i > 0} \left[\left(\left(w^T A\right)_i - q_i^g \right) x_i \right] \right\}.$$

For the infima inside we have for $i = 1, ..., n$,

$$\inf_{x_i > 0} \left[\left(\left(w^T A \right)_i - q_i^g \right) x_i \right] = \begin{cases} 0, & \left(w^T A \right)_i - q_i^g \geq 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Let us drag these results along the dual problem, that is now

$$(D_B) \qquad \sup_{\substack{q^f \in \mathbb{R}^n, w \in \mathbb{R}^m, q_i^g \geq e^{q_i^f - 1}, \\ \left(w^T A \right)_i - q_i^g \geq 0, i = 1, ..., n}} \left\{ \sum_{i=1}^n q_i^f - w^T b \right\},$$

rewritable as

$$(D_B) \qquad \sup_{\substack{q^g \in \operatorname{int}(\mathbb{R}_+^n), w \in \mathbb{R}^m, \\ \left(w^T A \right)_i - q_i^g \geq 0, i = 1, ..., n}} \left\{ -w^T b + \sum_{i=1}^n \sup_{q_i^f \leq 1 + \ln q_i^g} \left\{ q_i^f \right\} \right\}.$$

As the suprema after q_i^f, $i = 1, ..., n$, are trivially computable we get for the dual problem the following continuation

$$(D_B) \qquad \sup_{\substack{q^g \in \operatorname{int}(\mathbb{R}_+^n), w \in \mathbb{R}^m, \\ \left(w^T A \right)_i - q_i^g \geq 0, i = 1, ..., n}} \left\{ -w^T b + \sum_{i=1}^n \left(1 + \ln q_i^g \right) \right\}.$$

The variable q^g may also be retired, but in this case another constraint appears, namely $w^T A > 0$. For the sake of simplicity let us perform this step, too. The following problem is the ultimate dual problem to (P_B'')

$$(D_B) \qquad \sup_{\substack{w \in \mathbb{R}^m, \\ w^T A > 0}} \left\{ n - w^T b + \sum_{i=1}^n \ln \left(w^T A \right)_i \right\}.$$

Since the constraints of the primal problem (P_B) are linear and all feasible points x are in $\operatorname{int}(\mathbb{R}_+^n) = \operatorname{ri}(\mathbb{R}_+^n)$, no constraint qualification is required in this case. We can formulate the strong duality and optimality conditions statements right away. These assertions do not appear in the cited article, but we give them without proofs since these are similar to the ones already presented in the paper. There is a difference between the strong duality notion used here and the previous ones because normally we would present strong duality between (P_B'') and (D_B). But since the starting problem is (P_B) we modify a bit the statements using the obvious result $v(P_B) = -v(P_B'')$.

Theorem 13. *Provided that the primal problem (P_B) has at least a feasible point, the dual problem (D_B) has at least an optimal solution where it attains its* *maximal value and the sum of the optimal objective values of the two problems is vanishing, i.e.* $v(P_B) + v(D_B) = 0$.

Theorem 14. *(a) If the primal problem* (P_B) *has an optimal solution* $\bar{x}$*, then the dual problem* (D_B) *has also an optimal solution* $\bar{w}$ *and the following optimality conditions hold*

(i) $\ln \bar{x}_i + \ln \left(\bar{w}^T A\right)_i + n = \bar{w}^T A\bar{x}$,

(ii) $\bar{w}^T(A\bar{x} - b) = 0$.

(b) If $\bar{x}$ *is a feasible point to* (P_B) *and* $\bar{w}$ *is feasible to* (D_B) *such that the optimality conditions* (i) *and* (ii) *are true, then* $v(P_B)+v(D_B) = 0$*,* $\bar{x}$ *is an optimal solution to* (P_B) *and* $\bar{w}$ *an optimal solution to* (D_B)*.*

4 Conclusions and further ideas

We have considered an entropy-like optimization problem (P) whose dual problem has been calculated using a classical method through a special construction. Further the strong duality assertion and necessary and sufficient optimality conditions were presented. Three well-known entropy optimization problems picked from the literature were brought as applications to the problem we considered.

Getting back to the main problem, (P), an interesting question arises: why do functions f_i, $i = 1, ..., k$, have to be affine? Is the convexity not enough? We covered also this aspect. Considering the mentioned functions convex the method we used to derive a dual problem to (P) would have been utilizable only if the additional constraint $f_i(x) \geq g_i(x)$ $\forall x \in X$ such that $h(x) \geqq 0$, $i = 1, ..., k$, were posed. We considered also treating the so-modified problem, but applications to it appear too seldom. For instance the three special cases we treated could not be trapped into such a form without particularizing them more.

References

[1] Ben-Tal, A., Teboulle, M., Charnes, A. (1988): *The role of duality in optimization problems involving entropy functionals with applications to information theory.* Journal of Optimization Theory and Applications 58 (2), 209–223

[2] Boţ, R.I., Grad, S.M., Wanka, G. (2003): *Maximum entropy optimization for text classification problems.* In: W. Habenicht, B. Scheubrein, R. Scheubein (Eds.) - "Multi-Criteria- und Fuzzy-Systeme in Theorie und Praxis", Deutscher Universitäts-Verlag, Wiesbaden, 247–260

[3] Boţ, R.I., Grad, S.M., Wanka, G. (2004): *Entropy constrained programs and geometric duality obtained via Fenchel-Lagrange duality approach.* Submitted for publication.

[4] Censor, Y., De Pierro, A.R., Iusem, A.N. (1991): *Optimization of Burg's entropy over linear constraints.* Applied Numerical Mathematics 7(2), 151–165

[5] Censor, Y., Lent, A. (1987): *Optimization of "log x" entropy over linear equality constraints.* SIAM Journal on Control and Optimization 25(4), 921–933

[6] Elster, K.H., Reinhart, R., Schäuble, M., Donath, G. (1977): *Einführung in die nichtlineare Optimierung.* BSB B.G. Teubner Verlagsgesellschaft, Leipzig

[7] Fang, S.C., Rajasekera, J.R., Tsao, H.-S.J. (1997): *Entropy optimization and mathematical programming.* Kluwer Academic Publishers, Boston

[8] Golan, A., Dose, V. (2002): *Tomographic reconstruction from noisy data.* In "Bayesian inference and maximum entropy methods in science and engineering", AIP Conference Proceedings 617, 248–258

[9] Kapur, J.N. (1991): *Burg and Shannon's measures of entropy as limiting cases of two families of measures of entropy.* Proceedings of the National Academy of Sciences India 61A(3), 375–387

[10] Kapur, J.N., Kesavan, H.K. (1992): *Entropy optimization principles with applications.* Academic Press Inc., San Diego

[11] Noll, D. (1997): *Restoration of degraded images with maximum entropy.* Journal of Global Optimization 10(1) , 91–103

[12] Rockafellar, R.T. (1970): *Convex analysis.* Princeton University Press, Princeton

[13] Scott, C.H., Jefferson , T.R., Jorjani, S. (1995): *Conjugate duality in facility location.* In: Z. Drezner (Ed.) - "Facility Location: A Survey of Applications and Methods", Springer Verlag, 89–101

[14] Wanka, G., Boţ, R.I. (2002): *Multiobjective duality for convex ratios.* Journal of Mathematical Analysis and Applications 275 (1), 354–368