Exercises: Convex analysis

February 23, 2026

Convex sets

Exercise 1 (Perspective function). *Let $P : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be the* perspective function *defined as $P(x,t) = x/t$, with $\mathrm{dom}(P) = \mathbb{R}^n \times \mathbb{R}_+^*$.*

1. *Show that the image by P of the segment $[\binom{x}{s}, \binom{y}{t}]$ is the segment $[P(\binom{x}{s}), P(\binom{y}{t})]$, i.e. $P([\binom{x}{s}, \binom{y}{t}]) = [P(\binom{x}{s}), P(\binom{y}{t})]$.*
2. *Show that, if $C \subset \mathbb{R}^n \times \mathbb{R}_+^*$ is convex, then $P(C)$ is convex.*
3. *Show that, if $D \subset \mathbb{R}^n$, then $P^{-1}(D)$ is convex.*

Answers:

1. Let $\binom{x}{s} = (x,s)$ and $\binom{y}{t} = (y,t)$ be elements of $\mathbb{R}^n \times \mathbb{R}_+^*$.

$$P(\theta\tbinom{x}{s} + (1-\theta)\tbinom{y}{t}) = \frac{\theta x + (1-\theta)y}{\theta s + (1-\theta)t} = \mu\, P(\tbinom{x}{s}) + (1-\mu)\, P(\tbinom{y}{t}),$$

with $\mu(\theta) = \frac{\theta s}{\theta s + (1-\theta)t}$. Note that $\theta \mapsto \mu(\theta)$ is monotone and $\mu([0,1]) = [0,1]$. Thus, $P([\binom{x}{s}, \binom{y}{t}]) = [P(\binom{x}{s}), P(\binom{y}{t})]$.

2. Consider two elements of $P(C)$, $P(\binom{x}{s})$ and $P(\binom{y}{t})$. To show convexity we need to show that $[P(\binom{x}{s}), P(\binom{y}{t})] \subset P(C)$. By 1. we have $[P(\binom{x}{s}), P(\binom{y}{t})] = P([\binom{x}{s}, \binom{y}{t}])$ and $[\binom{x}{s}, \binom{y}{t}] \subset C$ by convexity of C.

3. Now assume that $\binom{x}{s} \in P^{-1}(D)$ and $\binom{y}{t} \in P^{-1}(D)$. We need to show that $\frac{\theta x + (1-\theta)y}{\theta s + (1-\theta)t} \in D$. This comes from $\frac{\theta x + (1-\theta)y}{\theta s + (1-\theta)t} = \mu(x/s) + (1-\mu)(y/t)$ with $\mu = \frac{\theta s}{\theta s + (1-\theta)t}$.

Exercise 2 (Dual cones)**.** *Recall that, for any set $K \subset \mathbb{R}^n$, $K^\oplus := \{y \in \mathbb{R}^n \mid \forall x \in K,\ \langle y, x\rangle \geq 0\}$. We say that K is self dual (which means that K is a closed convex cone) if $K^\oplus = K$.*

1. *Show that $K = \mathbb{R}_+^n$ is self dual.*
2. *We consider the set of symmetric matrices S_n with the scalar product $\langle A, B\rangle = \mathrm{tr}(AB)$. Show that $K = S_n^+(\mathbb{R})$ is self dual.*
3. *Let $\|\cdot\|$ be a norm, show that $K = \{(x,t) \mid \|x\| \leq t\}$ has for dual $K^\oplus = \{(z,\lambda) \mid \|z\|_\star \leq \lambda\}$, where $\|z\|_\star := \sup_{x:\|x\|\leq 1} z^\top x$.*

Answers:

1. obvious

2. Let $Y \in S_n \setminus S_n^+$. Then there exists $v \in \mathbb{R}^n$, $v^\top Y v < 0$. Moreover, $v^\top Y v = \mathrm{tr}(v^\top Y v) = \mathrm{tr}(vv^\top Y) < 0$. Hence we have $X = vv^\top \in S_n^+$ such that $\langle Y, X\rangle < 0$, i.e. $Y \notin (S_n^+)^\oplus$.

   On the other hand, consider $Y \in S_n^+$. We have the following decomposition $Y = \sum_{i=1}^n \lambda_i q_i^\top q_i$, where $\lambda_i \geq 0$ are the eigenvalues, and q_i the associated eigenvectors. Thus, for any $X \in S_n^+$, we have

$$\langle Y, X\rangle = \mathrm{tr}\Big(X \sum_{i=1}^n \lambda_i q_i q_i^\top\Big) = \mathrm{tr}\Big(\sum_{i=1}^n \lambda_i q_i^\top X q_i\Big) \geq 0$$

   hence $Y \in (S_n^+)^\oplus$.

3. We show the two inclusions.

   (i) If $(z,\lambda) \in K^\oplus$, then $\|z\|_\star \leq \lambda$. Take $(0,t) \in K$ for any $t \geq 0$. Then

$$0 \leq \langle (z,\lambda), (0,t)\rangle = \lambda t \quad \forall t \geq 0,$$

hence $\lambda \geq 0$. Now fix any u with $\|u\| \leq 1$ and any $t > 0$. The point $(x,t) = (-tu, t)$ belongs to K since $\|x\| = \|-tu\| \leq t$. Thus,

$$0 \leq \langle (z,\lambda), (-tu,t) \rangle = -t\, z^\top u + \lambda t = t(\lambda - z^\top u).$$

Dividing by $t > 0$ gives $z^\top u \leq \lambda$ for all $\|u\| \leq 1$. Taking the supremum over $\|u\| \leq 1$ yields $\|z\|_\star \leq \lambda$.

(ii) If $\|z\|_\star \leq \lambda$, then $(z,\lambda) \in K^\oplus$. Let $(x,t) \in K$, so $\|x\| \leq t$. By generalized Cauchy–Schwarz,

$$z^\top x \geq -|z^\top x| \geq -\|z\|_\star \|x\| \geq -\|z\|_\star t.$$

Therefore,

$$\langle (z,\lambda), (x,t) \rangle = z^\top x + \lambda t \geq (\lambda - \|z\|_\star)t \geq 0,$$

so $(z,\lambda) \in K^\oplus$.

Hence $K^\oplus = \{(z,\lambda) : \|z\|_\star \leq \lambda\}$.

Exercise 3 (Normal cones of standard convex sets)**.** *Compute the normal cone $N_C(x)$ for the following closed convex sets:*

1. *(Euclidean ball) $C = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$.*
2. *(Simplex) $\Delta = \{x \in \mathbb{R}^n : x \geq 0,\ \mathbf{1}^\top x = 1\}$.*

Bonus. *For the ball, compute the tangent cone $T_C(x)$ and verify $[T_C(x)]^\oplus = -N_C(x)$.*

Answers:

1. If $\|x\|_2 < 1$ (interior), $N_C(x) = \{0\}$. If $\|x\|_2 = 1$ (boundary), $N_C(x) = \{\lambda x : \lambda \geq 0\}$. (From supporting hyperplane of the ball: $x^\top y \leq 1$ at boundary point x.)

2. Let $x \in \Delta$ and define active set $I_0 = \{i : x_i = 0\}$. A vector $v \in N_\Delta(x)$ iff

   $$v = \mu\mathbf{1} - w, \qquad \mu \in \mathbb{R},\ w \geq 0, \text{ and } w_i = 0 \text{ for } i \notin I_0.$$

   Equivalently: $v_i = \mu$ on indices where $x_i > 0$, and $v_i \leq \mu$ where $x_i = 0$. (Reason: Δ is intersection of affine hyperplane $\{\mathbf{1}^\top x = 1\}$ and orthant $\{x \geq 0\}$, so the normal cone is sum of normals: span$(\mathbf{1})$ plus conic hull of $-e_i$ for active inequalities.)

Bonus: for the ball at $\|x\| = 1$, $T_C(x) = \{d : x^\top d \leq 0\}$ and polar gives $[T_C(x)]^\oplus = \{\lambda x : \lambda \leq 0\} = -N_C(x)$.

Convex functions

Exercise 4 (Recognizing convexity / strict / strong)**.** *For each function below, give: (i) its domain, (ii) whether it is convex, strictly convex, strongly convex on its domain (and provide a short justification).*
Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\lambda > 0$, and assume A has full column rank when stated.

1. *$f_1(x) = \|Ax - b\|_2$.*
2. *$f_2(x) = \frac{1}{2}\|Ax - b\|_2^2$.*
3. *$f_3(x) = \frac{1}{2}\|x\|_2^2 + \lambda\|x\|_1$.*
4. *$f_4(x) = -\log(b - a^\top x)$ with $\mathrm{dom}(f_4) = \{x : a^\top x < b\}$.*

Bonus. *For f_2, give a strong convexity modulus in terms of A (when A has full column rank).*

Answers:

1. Convex as composition of norm with affine map. Not strictly convex in general (e.g. if A not injective and/or norm not strictly convex along image). Not strongly convex.

2. Convex. If A has full column rank, then f_2 is strongly convex with modulus $\alpha = \sigma_{\min}(A)^2$ since $\nabla^2 f_2(x) = A^\top A \succeq \sigma_{\min}(A)^2 I$. Strict convexity holds when A injective.

3. Convex (sum of convex). Not differentiable. Not strongly convex from the ℓ_1 term, but the quadratic makes it strongly convex: modulus 1 w.r.t. $\|\cdot\|_2$.

4. Convex on its (open convex) domain since $-\log$ is convex and nondecreasing and composed with affine. Strictly convex because $-\log$ is strictly convex and $a \neq 0$. Not strongly convex on the full domain (Hessian blows up near boundary but no uniform lower bound on all of dom).

Exercise 5 (Moving average)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function.*

1. *Show that, $s \mapsto \int_0^1 f(st)dt$ is convex.*

2. *Show that, $\mathbb{R}_+^* \ni T \mapsto 1/T \int_0^T f(t)dt$ is convex.*

Answers:

1. Obvious from convexity of f and monotonicity of the integral.

2. Change of variable $u = t/T$.

Exercise 6 (Partial infimum)**.** *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ be a convex function and $C \subset \mathbb{R}^m$ a convex set. Show that the function*

$$g : x \mapsto \inf_{y \in C} f(x, y)$$

is convex.

Answers: Consider x_1 and x_2 in $\mathrm{dom}(g)$. For $\varepsilon > 0$, we have y_i such that $f(x_i, y_i) \leq g(x_i) + \varepsilon$. Thus,

$$\begin{aligned}
&g(\theta x_1 + (1-\theta)x_2) \\
&\quad = \inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \\
&\quad \leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\
&\quad \leq \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\
&\quad \leq \theta g(x_1) + (1-\theta) g(x_2) + \varepsilon.
\end{aligned}$$

Taking the limit in ε yields the result.

Exercise 7 (log determinant)**.** *Let, for any $X \in S_n$, $f(X) = \ln(\det(X))$ for $X \succ 0$, $-\infty$ otherwise. Consider, for $Z \succ 0$, and $V \in S_n$, the function $g : t \mapsto f(Z + tV)$.*

1. *Show that $g(t) = \sum_{i=1}^n \ln(1 + t\lambda_i) + f(Z)$, where the λ_i are the eigenvalues of $Z^{-1/2} V Z^{-1/2}$.*

2. *Show that g is concave. Conclude that f is concave.*

Answers:

1. We have

$$\begin{aligned}
g(t) &= f(Z^{1/2}(I + tZ^{-1/2} V Z^{-1/2}) Z^{1/2}) \\
&= \ln\det(Z) + \ln\det(I + tZ^{-1/2} V Z^{-1/2}) \\
&= f(Z) + \sum_{i=1}^n \ln(1 + t\lambda_i).
\end{aligned}$$

2. Concavity of g is obvious as sum of concave functions. We have $f(tX + (1-t)Y) = g(t)$, with $Z = X$ and $V = Y - X$. Hence f is concave.

Exercise 8 (Perspective function)**.** *Let $\phi : E \to \mathbb{R} \cup \{+\infty\}$. The perspective of ϕ is defined as $\tilde{\phi} : \mathbb{R}_+^* \times E \to \mathbb{R}$ by*

$$\tilde{\phi}(\eta, y) := \eta \phi(y/\eta).$$

Show that ϕ is convex iff $\tilde{\phi}$ is convex.

Answers:

$$\begin{aligned}
(\eta, y, z) \in \operatorname{epi} \tilde{\phi} &\Leftrightarrow \eta \phi(y/\eta) \leq z \\
&\Leftrightarrow \phi(y/\eta) \leq z/\eta \\
&\Leftrightarrow (y/\eta, z/\eta) \in \operatorname{epi} \phi.
\end{aligned}$$

Thus $\operatorname{epi} \phi$ is the image of $\operatorname{epi} \tilde{\phi}$ through the perspective function which preserves convexity (see Exercise 1).

Fenchel transform and subdifferential

Exercise 9 (Norm)**.** *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and $\|y\|_\star := \sup_{x : \|x\| \leq 1} y^\top x$ be its dual norm. Let $f : x \mapsto \|x\|$. Compute $f^\star$ and $\partial f(0)$.*

Answers: Recall that $f^\star(y) = \sup_x y^\top x - \|x\|$. We have $y^\top x \leq \|x\| \, \|y\|_\star$. Thus, if $\|y\|_\star \leq 1$, we have $f^\star(y) \leq \sup_x \|x\|(\|y\|_\star - 1) \leq 0$, attained for $x = 0$.
Otherwise, if $\|y\|_\star > 1$, there exists x such that $y^\top x > \|x\|$, and for all $t > 0$, $f^\star(y) \geq t(y^\top x - \|x\|)$, hence $f^\star(y) = +\infty$. Consequently $f^\star(y) = \mathbb{I}_{\{\|y\|_\star \leq 1\}}$.
By Fenchel–Young, $\partial f(0) = \{y \in \mathbb{R}^n \mid \|y\|_\star \leq 1\}$.

Exercise 10 (Lasso)**.** *Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\lambda > 0$ and consider*

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_1.$$

1. *Show that the problem admits at least one solution, and is unique if A has full column rank.*

2. *Compute $\partial \|x\|_1$ and derive the optimality condition for a minimizer $x^\sharp$:*

$$0 \in A^\top (Ax^\sharp - b) + \lambda \, \partial \|x^\sharp\|_1.$$

3. *Prove the coordinate-wise characterization:*

$$x_i^\sharp \neq 0 \;\Rightarrow\; a_i^\top(Ax^\sharp - b) = -\lambda\,\mathrm{sign}(x_i^\sharp)$$

$$x_i^\sharp = 0 \;\Rightarrow\; |a_i^\top(Ax^\sharp - b)| \leq \lambda,$$

where a_i is the i-th column of A.

4. *If $\lambda \geq \|A^\top b\|_\infty$, then $x^\sharp = 0$ is optimal.*

5. *Interpretation: explain in one sentence why large λ promotes sparsity of $x^\sharp$.*

Answers:

1. **Existence and uniqueness.** Define

$$F(x) := \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1.$$

F is proper and lower semicontinuous (sum of continuous and lsc convex functions). Moreover F is coercive since $\|x\|_1 \to +\infty$ as $\|x\|_2 \to +\infty$, hence $F(x) \to +\infty$ and a minimizer exists.

If A has full column rank then $A^\top A \succ 0$ and $x \mapsto \frac{1}{2}\|Ax - b\|^2$ is strongly convex, e.g. with modulus $\sigma_{\min}(A)^2$ since

$$\nabla^2\Big(\tfrac{1}{2}\|Ax - b\|^2\Big) = A^\top A \succeq \sigma_{\min}(A)^2 I.$$

Adding the convex term $\lambda\|x\|_1$ preserves strong convexity, so the minimizer is unique.

2. **Subdifferential of $\|\cdot\|_1$ and optimality condition.**

First in dimension 1, for $h(t) = |t|$,

$$\partial h(t) = \begin{cases} \{1\}, & t > 0, \\ [-1, 1], & t = 0, \\ \{-1\}, & t < 0. \end{cases}$$

Justification: $g \in \partial h(t)$ iff for all $s \in \mathbb{R}$,

$$|s| \geq |t| + g(s - t).$$

If $t > 0$, take $s = t + \varepsilon$ and $s = t - \varepsilon$, then $g = 1$. If $t < 0$, similarly $g = -1$. If $t = 0$, the condition becomes $|s| \geq gs$ for all s, which holds iff $g \in [-1, 1]$.

Now $\|x\|_1 = \sum_{i=1}^n |x_i|$. Each term depends on a single coordinate, hence

$$\partial\|x\|_1 = \Big\{ s \in \mathbb{R}^n : \; s_i \in \partial|x_i| \;\; \forall i \in [n] \Big\}.$$

Therefore, for a minimizer $x^\sharp$, the convex optimality condition gives

$$0 \in \nabla\Big(\tfrac{1}{2}\|Ax - b\|_2^2\Big)\Big|_{x = x^\sharp} + \lambda\,\partial\|x^\sharp\|_1$$

that is

$$0 \in A^\top(Ax^\sharp - b) + \lambda\,\partial\|x^\sharp\|_1.$$

3. **Coordinate-wise characterization.** The inclusion means that there exists $s \in \partial\|x^\sharp\|_1$ such that

$$A^\top(Ax^\sharp - b) + \lambda s = 0.$$

Taking the i-th coordinate (with a_i the i-th column of A) yields

$$a_i^\top(Ax^\sharp - b) + \lambda s_i = 0, \qquad s_i \in \partial|x_i^\sharp|.$$

Hence

$$a_i^\top(Ax^\sharp - b) = \begin{cases} -\lambda\,\mathrm{sign}(x_i^\sharp), & x_i^\sharp \neq 0, \\ \in [-\lambda, \lambda], & x_i^\sharp = 0. \end{cases}$$

Equivalently,

$$\begin{cases} x_i^\sharp \neq 0 \Rightarrow a_i^\top(Ax^\sharp - b) = -\lambda\,\mathrm{sign}(x_i^\sharp), \\ x_i^\sharp = 0 \Rightarrow |a_i^\top(Ax^\sharp - b)| \leq \lambda. \end{cases}$$

4. **If $\lambda \geq \|A^\top b\|_\infty$, then $x^\sharp = 0$ is optimal.**

Check the optimality condition at $x = 0$. We have

$$\nabla\Big(\tfrac{1}{2}\|Ax - b\|^2\Big)\Big|_{x=0} = A^\top(A \cdot 0 - b) = -A^\top b.$$

So $x^\sharp = 0$ is optimal iff

$$0 \in -A^\top b + \lambda\,\partial\|0\|_1.$$

But $\partial\|0\|_1 = [-1, 1]^n$, hence the inclusion is equivalent to the existence of $s \in [-1, 1]^n$ with

$$-A^\top b + \lambda s = 0 \quad \Longleftrightarrow \quad s = \frac{A^\top b}{\lambda} \in [-1, 1]^n.$$

This holds iff

$$\|A^\top b\|_\infty \leq \lambda,$$

which proves the claim.

5. **Interpretation (sparsity).** A coordinate can be set to zero whenever

$$|a_i^\top (Ax^\sharp - b)| \le \lambda.$$

Increasing λ makes these inequalities easier to satisfy, so more coordinates become zero.

Exercise 11 (Fenchel calculus: indicator, support, and affine change)**.** *Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set and define the indicator $\mathbb{I}_C$ and the support function*

$$\sigma_C(y) := \sup_{x \in C} y^\top x.$$

1. *Show that $(\mathbb{I}_C)^\star = \sigma_C$.*

2. *Compute σ_C for:*

   (a) $C = B_2(0,1) = \{x : \|x\|_2 \le 1\}$,

   (b) $C = \{x : \|x\|_\infty \le 1\}$.

3. *Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and define $f(x) = \mathbb{I}_C(Ax + b)$. Give an expression for $f^\star$ (you may state the formula with the condition under which it holds).*

Answers:

1. By definition: $(\mathbb{I}_C)^\star(y) = \sup_x\{y^\top x - \mathbb{I}_C(x)\} = \sup_{x\in C} y^\top x = \sigma_C(y)$.

2. (a) $\sigma_{B_2}(y) = \sup_{\|x\|\le 1} y^\top x = \|y\|_2$. (b) $\sigma_{\{\|x\|_\infty \le 1\}}(y) = \sup_{\|x\|_\infty\le 1} \sum_i y_i x_i = \sum_i |y_i| = \|y\|_1$.

3. One standard form: if $\mathrm{Im}(A) \cap \mathrm{ri}(\mathrm{dom}\, \mathbb{I}_C) = \mathrm{Im}(A) \cap \mathrm{ri}(C) \neq \emptyset$, then

$$f^\star(u) = \sigma_C(A^\top u) - b^\top u.$$

(Indeed $f = \delta_C(Ax + b)$ is composition with affine map; conjugate is support with $A^\top u$ and shift gives $-b^\top u$.)

Exercise 12 (Log sum exp)**.** *We consider $f(x) := \ln(\sum_{i=1}^n e^{x_i})$.*

1. *Show that f is convex. Hint : recall Holder's inequality $x^\top y \le \|x\|_p \|y\|_q$ for $1/p + 1/q = 1$.*

2. *Show that $f^\star(y) = \sum_{i=1}^n y_i \ln(y_i)$ if $y \ge 0$ and $\sum_i y_i = 1$, $+\infty$ otherwise.*

Answers:

1. Let $x, y \in \mathbb{R}^n$ and set $u_i = e^{x_i}$ and $v_i = e^{y_i}$. For $\theta \in [0,1]$, we have

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \ln\Big(\sum_{i=1}^n e^{\theta x_i + (1-\theta) y_i}\Big) \\ &= \ln\Big(\sum_{i=1}^n u_i^\theta v_i^{1-\theta}\Big). \end{aligned}$$

We use $p = 1/\theta$ and $q = 1/(1-\theta)$ in Hölder's inequality to get

$$\begin{aligned} f(\theta x + (1-\theta)y) &\le \ln\left(\left(\sum_{i=1}^n u_i\right)^\theta \left(\sum_{i=1}^n v_i\right)^{1-\theta}\right) \\ &= \theta f(x) + (1-\theta) f(y). \end{aligned}$$

2. We compute the Fenchel conjugate

$$f^\star(y) = \sup_{x\in\mathbb{R}^n} \Big\{y^\top x - \log\Big(\sum_{i=1}^n e^{x_i}\Big)\Big\}.$$

Step 1: the domain constraints.

- If $\sum_{i=1}^n y_i \neq 1$, then the supremum is $+\infty$. Indeed, for any $c \in \mathbb{R}$ let $x' = x + c\mathbf{1}$. Then

$$y^\top x' - \log\Big(\sum_i e^{x_i'}\Big) = \big(y^\top x - \log \sum_i e^{x_i}\big) + c\Big(\sum_i y_i - 1\Big).$$

  Sending $c \to +\infty$ or $c \to -\infty$ yields $+\infty$ whenever $\sum_i y_i \neq 1$.

- If $y_j < 0$ for some j, then the supremum is $+\infty$. Take $x = te_j$ with $t \to -\infty$. Then

$$y^\top x - \log\Big(\sum_i e^{x_i}\Big) = y_j t - \log\big(e^t + (n-1)\big) \sim y_j t - \log(n-1$$

  since $y_j < 0$.

Therefore $f^\star(y) = +\infty$ unless $y \in \Delta := \{y \in \mathbb{R}^n : \ y \ge 0, \ \sum_i y_i = 1\}$.

Step 2: maximize over x for $y \in \Delta$. For $y \in \Delta$, the objective

$$\Phi(x) := y^\top x - \log\left(\sum_{i=1}^{n} e^{x_i}\right)$$

is concave in x (linear minus convex), so any critical point is a global maximizer. Compute the gradient:

$$\frac{\partial \Phi}{\partial x_i}(x) = y_i - \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}.$$

At an optimum $x^\star$ we must have

$$y_i = \frac{e^{x_i^\star}}{\sum_{j=1}^{n} e^{x_j^\star}} \quad \Longleftrightarrow \quad x_i^\star = \log y_i + c,$$

for some constant $c \in \mathbb{R}$ (when $y_i = 0$, this corresponds to $x_i^\star \to -\infty$; the value below remains valid with the convention $0 \log 0 = 0$).

Plugging $x^\star = \log y + c\mathbf{1}$ into Φ:

$$\sum_i e^{x_i^\star} = e^c \sum_i y_i = e^c, \qquad y^\top x^\star = \sum_i y_i(\log y_i + c) = \sum_i y_i \log y_i + c,$$

hence

$$\Phi(x^\star) = \left(\sum_i y_i \log y_i + c\right) - \log(e^c) = \sum_{i=1}^{n} y_i \log y_i.$$

Conclusion.

$$f^\star(y) = \begin{cases} \displaystyle\sum_{i=1}^{n} y_i \log y_i & \text{if } y \geq 0,\ \sum_{i=1}^{n} y_i = 1, \\ +\infty & \text{otherwise,} \end{cases}$$

with the convention $0 \log 0 = 0$.