Maximum Entropy on the Mean and the Cramér Rate Function in Statistical Estimation: Properties, Models and Algorithms
– Supplemental Material –

April 6, 2023

1 Bregman Proximal Operators

In this section we present a detailed derivation of the Bregman proximal operators. We recall that $\Delta_d := \{y \in \mathbb{R}^d_+ : \langle e, y\rangle = 1\}$ with $e = (1, 1, \ldots , 1)^T \in \mathbb{R}^d$ stands for the unit simplex and that for any $p \in \mathbb{R}^d$ we denote $I(p) := \{y \in \mathbb{R}^d : y_i = 0 \ (p_i = 0)\}$ and $\Delta_{(d)} := \{y \in \mathbb{R}^d_+ : \langle e, y\rangle \leq 1\}$.

1.1 Smooth Adaptable Kernel - Energy (Data Fidelity Reference Distribution - Normal)

In this case, the Bregman proximal operator reduces to the classical proximal operator:

$$x^+ = \operatorname{argmin}\left\{\psi^*_R(u) + \frac{1}{2t}\|u - \bar{x}\|_2^2 : u \in \mathbb{R}^d\right\}.$$

Multivariate Normal

The proximal gradient operator is given by:

$$\begin{aligned} x^+ &= \operatorname{argmin}\left\{\psi^*_R(u) + \frac{1}{2t}\|u - \bar{x}\|_2^2 : u \in \mathbb{R}^d\right\} \\ &= \operatorname{argmin}\left\{\frac{1}{2}(u-\mu)^T\Sigma^{-1}(u-\mu) + \frac{1}{2t}\|u - \bar{x}\|_2^2 : u \in \mathbb{R}^d\right\}. \end{aligned}$$

The first-order optimality condition imply

$$\Sigma^{-1}(x^+ - \mu) + (x^+ - \bar{x})/t = 0 \quad \Rightarrow \quad x^+ = (tI + \Sigma)^{-1}(\Sigma\bar{x} + t\mu).$$

Multivariate Normal-inverse Gaussian

The proximal gradient operator is given by:

$$\begin{aligned} x^+ &= \operatorname{argmin}\left\{\psi^*_R(u) + \frac{1}{2t}\|u - \bar{x}\|^2 : u \in \mathbb{R}^d\right\} \\ &= \operatorname{argmin}\left\{\alpha\sqrt{\delta^2 + (u-\mu)^T\Sigma^{-1}(u-\mu)} - \langle\beta, u-\mu\rangle - \delta\gamma + \frac{1}{2t}\|u - \bar{x}\|^2 : u \in \mathbb{R}^d\right\}. \end{aligned}$$

Using the first order optimality conditions we obtain that

$$x^+ + \frac{\alpha t \Sigma^{-1}(x^+ - \mu)}{\sqrt{\delta^2 + (x^+ - \mu)^T \Sigma^{-1}(x^+ - \mu)}} = t\beta + \bar{x} \quad \Rightarrow \quad x^+ = \left(I + \rho\Sigma^{-1}\right)^{-1}\left(t\beta + \bar{x} + \rho\Sigma^{-1}\mu\right),$$

where

$$\rho := \frac{\alpha t}{\sqrt{\delta^2 + (x^+ - \mu)^T \Sigma^{-1}(x^+ - \mu)}} \geq 0.$$

Thus

$$\begin{aligned} x^+ - \mu &= \left(I + \rho\Sigma^{-1}\right)^{-1}\left(t\beta + \bar{x} + \rho\Sigma^{-1}\mu - \left(I + \rho\Sigma^{-1}\right)\mu\right) \\ &= \left(I + \rho\Sigma^{-1}\right)^{-1}\left(t\beta + \bar{x} - \mu\right), \end{aligned}$$

and we can summarize that

$$x^+ = \left(I + \rho\Sigma^{-1}\right)^{-1}\left(t\beta + \bar{x} + \rho\Sigma^{-1}\mu\right),$$

where (using the notation $\|z\|_M^2 = \langle z, Mz\rangle$ for $z \in \mathbb{R}^d$ and $M \in \mathbb{R}^{d\times d}$ such that $M \succ 0$)

$$\rho \in \mathbb{R}_+ : \quad (\rho\delta)^2 + \|\left(\rho^{-1}I + \Sigma^{-1}\right)^{-1}\left(t\beta + \bar{x} - \mu\right)\|_{\Sigma^{-1}}^2 = (\alpha t)^2.$$

Gamma

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, we would like to solve

$$\begin{aligned} x^+ &= \operatorname{argmin}\left\{\psi_R^*(u) + \frac{1}{2t}(u - \bar{x})^2 : u \in \mathbb{R}\right\} \\ &= \operatorname{argmin}\left\{\beta u - \alpha + \alpha\log\left(\frac{\alpha}{\beta u}\right) + \frac{1}{2t}(u - \bar{x})^2 : u \in \mathbb{R}_{++}\right\}. \end{aligned}$$

By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$. Thus, x^+ must satisfy the first-order optimality condition

$$\beta - \alpha/x^+ + (x^+ - \bar{x})/t = 0 \quad \Rightarrow \quad (x^+)^2 - (\bar{x} - t\beta)x^+ - t\alpha = 0.$$

Thus, x^+ is the positive root of the quadratic equation given above

$$x^+ = \frac{\bar{x} - t\beta + \sqrt{(\bar{x} - t\beta)^2 + 4t\alpha}}{2}.$$

Laplace

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, we would like to solve

$$x^+ = \operatorname{argmin}\left\{\psi_R^*(u) + \frac{1}{2t}(u - \bar{x})^2 : u \in \mathbb{R}\right\},$$

where

$$\psi_R^*(u) := \begin{cases} 0, & u = \mu, \\ \sqrt{1 + \left(\frac{u-\mu}{b}\right)^2} - 1 + \log\left(2\left(\frac{u-\mu}{b}\right)^{-2}\left[\sqrt{1+\left(\frac{u-\mu}{b}\right)^2} - 1\right]\right), & u \neq \mu. \end{cases}$$

Denoting $\rho = (u-\mu)/b$, we obtain that $x^+ = \mu + b\rho^+$ where ρ^+ is given by

$$\rho^+ \quad = \operatorname{argmin}\left\{\tfrac{1}{2}(b\rho + \mu - \bar{x})^2 + t\psi_R^*(\mu + b\rho) : \rho \in \mathbb{R}\right\}.$$

In view of Corollary 5.1 we know that if $\bar{x} = E_R = \mu$ then $x^+ = \bar{x} = \mu$ and $\rho^+ = 0$. In order to evaluate the value of ρ^+, and as a result of x^+, when $\bar{x} \neq \mu$ we can consider the first-order optimality conditions. We denote $\xi(\rho) := \frac{1}{2}(b\rho + \mu - \bar{x})^2$ and write the optimality conditions as follows

$$\begin{aligned} & \xi'(\rho) + t\left[\frac{\rho}{\sqrt{1+\rho^2}} - \frac{2}{\rho} + \frac{1}{\sqrt{1+\rho^2}-1}\frac{\rho}{\sqrt{1+\rho^2}}\right] = 0 \\ \Leftrightarrow \quad & \xi'(\rho) + t\left[\frac{\rho}{\sqrt{1+\rho^2}}\left(1 + \frac{1}{\sqrt{1+\rho^2}-1}\right) - \frac{2}{\rho}\right] = 0 \\ \Leftrightarrow \quad & \xi'(\rho) + t\left[\frac{\rho}{\sqrt{1+\rho^2}-1} - \frac{2}{\rho}\right] = 0 \\ \Leftrightarrow \quad & \frac{t\rho}{\sqrt{1+\rho^2}-1} = \frac{2t}{\rho} - \xi'(\rho) = \frac{\vartheta(\rho)}{\rho} \end{aligned}$$

where $\vartheta(\rho) := 2t - \rho\xi'(\rho)$. The above expression is the same as

$$\begin{aligned} & \vartheta(\rho)\sqrt{1+\rho^2} = t\rho^2 + \vartheta(\rho) \\ \Leftrightarrow \quad & \vartheta(\rho)^2(1+\rho^2) = (t\rho^2 + \vartheta(\rho))^2 = t^2\rho^4 + 2t\rho^2\vartheta(\rho) + \vartheta(\rho)^2, \end{aligned}$$

Thus, using the assumption that $\rho \neq 0$ we obtain that

$$\begin{aligned} & \vartheta(\rho)^2(1+\rho^2) = t^2\rho^4 + 2t\rho^2\vartheta(\rho) + \vartheta(\rho)^2 \\ \Leftrightarrow \quad & \vartheta(\rho)^2 = t^2\rho^2 + 2t\vartheta(\rho) \\ \Leftrightarrow \quad & (\vartheta(\rho)/t - 1)^2 = \rho^2 + 1. \end{aligned} \tag{1.1}$$

Plugging $\vartheta(\rho) := 2t - \rho\xi'(\rho) = 2t - b\rho(b\rho + \mu - \bar{x})$ we can write the above as

$$\begin{aligned}
& & \left(1 - (b/t)b\rho^2 - (b/t)(\mu - \bar{x})\rho\right)^2 = \rho^2 + 1 \\
&\Leftrightarrow & 1 - 2\left((b/t)b\rho^2 + (b/t)(\mu - \bar{x})\rho\right) + \left((b/t)b\rho^2 + (b/t)(\mu - \bar{x})\rho\right)^2 = \rho^2 + 1 \\
&\Leftrightarrow & (b/t)^2\left(b\rho^2 + (\mu - \bar{x})\rho\right)^2 - 2(b/t)\left(b\rho^2 + (\mu - \bar{x})\rho\right) = \rho^2 \\
&\Leftrightarrow & (b/t)^2\rho\left(b\rho + (\mu - \bar{x})\right)^2 - 2(b/t)\left(b\rho + (\mu - \bar{x})\right) = \rho \\
&\Leftrightarrow & (b/t)^2\rho\left(b^2\rho^2 + (\mu - \bar{x})^2 + 2(\mu - \bar{x})b\rho\right) - 2(b/t)\left(b\rho + (\mu - \bar{x})\right) = \rho \\
&\Leftrightarrow & (b/t)^2b^2\rho^3 + 2(b/t)^2b(\mu - \bar{x})\rho^2 + [(b/t)^2(\mu - \bar{x})^2 - 2(b/t)b - 1]\rho - 2(b/t)(\mu - \bar{x}) = 0 \\
&\Leftrightarrow & \alpha_1\rho^3 + \alpha_2\rho^2 + \alpha_3\rho + \alpha_4 = 0,
\end{aligned}$$

where $\alpha = [(b/t)^2b^2, 2(b/t)^2b(\mu - \bar{x}), (b/t)^2(\mu - \bar{x})^2 - 2(b/t)b - 1, -2(b/t)(\mu - \bar{x})]^T$.
To summarize,

$$x^+ = \begin{cases} \mu, & \mu = \bar{x}, \\ \mu + b\rho^+, & \mu \neq \bar{x}, \end{cases}$$

where ρ^+ is the unique real root of the cubic equation $\alpha_1\rho^3 + \alpha_2\rho^2 + \alpha_3\rho + \alpha_4 = 0$ with coefficients:

$$\begin{array}{ll} \alpha_1 = (b/t)^2b^2, & \alpha_3 = (b/t)^2(\mu - \bar{x})^2 - 2(b/t)b - 1 \\ \alpha_2 = 2(b/t)^2b(\mu - \bar{x}), & \alpha_4 = -2(b/t)(\mu - \bar{x}). \end{array}$$

Poisson

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, we would like to solve

$$\begin{aligned}
x^+ &= \operatorname{argmin}\left\{\psi_R^*(u) + \frac{1}{2t}(u - \bar{x})^2 : u \in \mathbb{R}\right\} \\
&= \operatorname{argmin}\left\{u\log(u/\lambda) - u + \lambda + \frac{1}{2t}(u - \bar{x})^2 : u \in \mathbb{R}_+\right\}.
\end{aligned}$$

By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$. Thus, x^+ must satisfy the first-order optimality condition

$$\log(x^+/\lambda) + (x^+ - \bar{x})/t = 0.$$

Multinomial

Let $p \in \Delta_{(d)}$ such that $\langle e, p\rangle < 1$. The proximal gradient step is given by:

$$\begin{aligned}
x^+ &= \operatorname{argmin}\left\{\frac{1}{2t}\|u - \bar{x}\|_2^2 + \psi_R^*(u) : u \in \mathbb{R}^d\right\} \\
&= \operatorname{argmin}\left\{\frac{1}{2t}\|u - \bar{x}\|_2^2 + \sum_{i=1}^d y_i\log\left(\frac{y_i}{np_i}\right) + (n - \langle e, y\rangle)\log\left(\frac{n - \langle e, y\rangle}{n(1 - \langle e, p\rangle)}\right) : u \in n\Delta_{(d)} \cap I(p)\right\}. \blacksquare
\end{aligned}$$

Evidently, $x_i^+ = 0$ for any $i \in \{1, 2, \ldots, d\}$ such that $p_i = 0$. We can write the Lagrangian

$$L(u; \lambda) = \frac{1}{2t}\|u - \bar{x}\|_2^2 + \sum_{i=1}^{d} y_i \log\left(\frac{y_i}{np_i}\right) + (n - \langle e, y\rangle)\log\left(\frac{n - \langle e, y\rangle}{n(1 - \langle e, p\rangle)}\right) + \lambda(\langle e, u\rangle - n),$$

where $\lambda \geq 0$ is the Lagrange multiplier. By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$. Thus, writing the KKT conditions, and noting that the complementary slackness condition and $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$ yields $\lambda = 0$, we obtain that the unique solution is

$$x^+ \in \mathbb{R}_+^d \cap I(p) : \quad (x_i^+ - \bar{x}_i)/t + \log\left(\frac{x_i^+(1 - \sum_{j=1}^d p_j)}{p_i(n - \sum_{j=1}^d x_j^+)}\right) = 0, \qquad i = 1, 2, \ldots, d.$$

Negative Multinomial

Consider the negative multinomial reference distribution with parameters $p \in \operatorname{ri}\Delta_d$, $x_0 \in \mathbb{R}_{++}$ and set $p_0 := 1 - \sum_{i=1}^d p_i$. The proximal gradient step is given by:

$$\begin{aligned} x^+ &= \operatorname{argmin}\left\{\psi_R^*(u) + \tfrac{1}{2t}\|u - \bar{x}\|_2^2 : u \in \mathbb{R}^d\right\} \\ &= \operatorname{argmin}\left\{\tfrac{1}{2t}\|u - \bar{x}\|_2^2 + \textstyle\sum_{i=1}^d u_i \log\left(\frac{u_i}{p_i(x_0 + \sum_{j=1}^d u_j)}\right) + x_0 \log\left(\frac{x_0}{x_0 + \sum_{j=1}^d u_j}\right) : u \in \mathbb{R}_+^d \cap I(p)\right\}. \end{aligned}$$

Evidently, $x_i^+ = 0$ for any $i \in \{1, 2, \ldots, d\}$ such that $p_i = 0$. By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$. Thus, $x^+ \in \mathbb{R}_+^d \cap I(p)$ must satisfy the first-order optimality conditions, i.e., for all $i \in \{j \in \{1, 2, \ldots, d\} : p_j \neq 0\}$

$$(x_i^+ - \bar{x}_i)/t + \log\left(\frac{x_i^+}{p_i(x_0 + \sum_{j=1}^d x_j^+)}\right) = 0.$$

Discrete Uniform

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, we would like to solve

$$x^+ = \operatorname{argmin}\left\{\psi_R^*(u) + \frac{1}{2t}(u - \bar{x})^2 : u \in \mathbb{R}\right\}.$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use the extended Moreau decomposition [1, Theorem 6.45] together with the fact that $\psi_P^{**} = \psi_P$ [1, Theorem 4.8]. Hence

$$x^+ = \operatorname{prox}_{t\psi_R^*}(\bar{x}) = \bar{x} - t\operatorname{prox}_{t^{-1}\psi_P}(\bar{x}/t).$$

Thus, we need to find

$$\begin{aligned} \theta^+ = \operatorname{prox}_{t^{-1}\psi_P}(\bar{x}/t) &= \operatorname{argmin}\left\{\psi_P(\theta) + \frac{t}{2}(\theta - \bar{x}/t)^2 : \theta \in \mathbb{R}\right\} \\ &= \operatorname{argmin}\left\{\log\left(M_R[\theta]\right) + \frac{t}{2}(\theta - \bar{x}/t)^2 : \theta \in \mathbb{R}\right\}, \end{aligned}$$

where we recall that

$$M_R[\theta] = \begin{cases} \frac{\exp((b+1)\theta) - \exp(a\theta)}{n(\exp(\theta)-1)}, & \theta \neq 0, \\ 1, & \theta = 0. \end{cases}$$

In view of Corollary 5.1 we know that if $\bar{x} = E_R = (a+b)/2$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq (b-a)/2$ we can consider the first-order optimality conditions

$$t(\theta^+ - \bar{x}/t) + \frac{(b+1)\exp((b+1)\theta^+) - a\exp(a\theta^+)}{\exp((b+1)\theta^+) - \exp(a\theta^+)} = \frac{\exp(\theta^+)}{\exp(\theta^+) - 1}.$$

To summarize, $x^+ = \bar{x} - t\theta^+$ where θ^+ is equal to zero if $\bar{x} = 0$, otherwise θ^+ is the solution to the above equation.

Continuous Uniform

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, we would like to solve

$$x^+ \quad = \operatorname{argmin} \left\{ \psi_R^*(u) + \frac{1}{2t}(u - \bar{x})^2 : u \in \mathbb{R} \right\}.$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use the extended Moreau decomposition [1, Theorem 6.45] together with the fact that $\psi_P^{**} = \psi_P$ [1, Theorem 4.8]. Hence

$$x^+ = \operatorname{prox}_{t\psi_R^*}(\bar{x}) = \bar{x} - t\operatorname{prox}_{t^{-1}\psi_P}(\bar{x}/t).$$

Thus, we need to find

$$\begin{aligned} \theta^+ = \operatorname{prox}_{t^{-1}\psi_P}(\bar{x}/t) \quad &= \operatorname{argmin} \left\{ \psi_P(\theta) + \frac{t}{2}(\theta - \bar{x}/t)^2 : \theta \in \mathbb{R} \right\} \\ &= \operatorname{argmin} \left\{ \log\left(M_R[\theta]\right) + \frac{t}{2}(\theta - \bar{x}/t)^2 : \theta \in \mathbb{R} \right\}, \end{aligned}$$

where we recall that

$$M_R[\theta] = \begin{cases} \frac{\exp(b\theta) - \exp(a\theta)}{\theta(b-a)}, & \theta \neq 0, \\ 1, & \theta = 0. \end{cases}$$

In view of Corollary 5.1 we know that if $\bar{x} = E_R = (a+b)/2$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq (b-a)/2$ we can consider the first-order optimality conditions

$$t(\theta^+ - \bar{x}/t) + \frac{b\exp(b\theta^+) - a\exp(a\theta^+)}{\exp(b\theta^+) - \exp(a\theta^+)} = \frac{1}{\theta^+}.$$

To summarize, $x^+ = \bar{x} - t\theta^+$ where θ^+ is equal to zero if $\bar{x} = 0$, otherwise θ^+ is the solution to the above equation.

Logistic

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, we would like to solve

$$x^+ = \operatorname{argmin}\left\{\psi_R^*(u) + \frac{1}{2t}(u - \bar{x})^2 : u \in \mathbb{R}\right\}.$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use the extended Moreau decomposition [1, Theorem 6.45] together with the fact that $\psi_P^{**} = \psi_P$ [1, Theorem 4.8]. Hence

$$x^+ = \operatorname{prox}_{t\psi_R^*}(\bar{x}) = \bar{x} - t\operatorname{prox}_{t^{-1}\psi_P}(\bar{x}/t).$$

Thus, we need to find

$$\begin{aligned}
\theta^+ = \operatorname{prox}_{t^{-1}\psi_P}(\bar{x}/t) \quad &= \operatorname{argmin}\left\{\psi_P(\theta) + \frac{t}{2}(\theta - \bar{x}/t)^2 : \theta \in \mathbb{R}\right\} \\
&= \operatorname{argmin}\left\{\log\left(M_R[\theta]\right) + \frac{t}{2}(\theta - \bar{x}/t)^2 : \theta \in \mathbb{R}\right\} \\
&= \operatorname{argmin}\left\{\log\left(\exp(\mu\theta)\phi_s(\theta)\right) + \frac{t}{2}(\theta - \bar{x}/t)^2 : \theta \in (-1/s, 1/s)\right\} \\
&= \operatorname{argmin}\left\{\log\left(\phi_s(\theta)\right) + \frac{t}{2}(\theta - (\bar{x} - \mu)/t)^2 : \theta \in (-1/s, 1/s)\right\},
\end{aligned}$$

where we recall that $\phi_s(\theta) = B(1 - s\theta, 1 + s\theta)$. In view of Corollary 5.1 we know that if $\bar{x} = E_R = \mu$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq \mu$, under similar arguments to those we mentioned when evaluated the Cramér rate function, we can write the first order optimality conditions of the above problem

$$0 = t\theta^+ - (\bar{x} - \mu) + \frac{\phi_s'(\theta^+)}{\phi_s(\theta^+)} \qquad \Rightarrow \qquad t\theta^+ + \frac{1}{\theta^+} + \frac{\pi s}{\tan\left(-\pi s\theta^+\right)} = \bar{x} - \mu,$$

and then the proximal operator yields the point $x^+ = \bar{x} - t\theta^+$.

1.2 Smooth Adaptable Kernel - Boltzmann-Shannon Entropy (Data Fidelity Reference Distribution - Poisson)

Omitting some constant terms we can write the Bregman proximal gradient operator as:

$$\begin{aligned}
x^+ \quad &= \operatorname{argmin}\left\{t\psi_R^*(u) + D_h(u, \bar{x}) : u \in \mathbb{R}^d\right\} \\
&= \operatorname{argmin}\left\{t\psi_R^*(u) + h(u) - \langle \nabla h(\bar{x}), u\rangle : u \in \mathbb{R}^d\right\},
\end{aligned} \tag{1.2}$$

where $h(x) = \sum_{i=1}^d x_i \log x_i$ and $\bar{z}_i := [\nabla h(\bar{x})]_i = \log(x_i) + 1$ for all $i = 1, 2, \ldots, d$. Namely, throughout we will be solving

$$x^+ = \operatorname{argmin}\left\{t\psi_R^*(u) + \sum_{i=1}^d u_i\left(\log\left(\frac{u_i}{\bar{x}_i}\right) - 1\right)\right\}.$$

Multivariate Normal

The multivariate Gaussian distribution admits parameters $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{S}^n_{++}$. We seek to find

$$\operatorname{argmin}\left\{\frac{t}{2}(u-\mu)^T\Sigma^{-1}(u-\mu)+\sum_{i=1}^{d} u_i\left(\log\left(\frac{u_i}{\bar{x}_i}\right)-1\right) : u \in \mathbb{R}^d\right\}$$

Denoting $X^+ = \operatorname{diag}(x^+) \succ 0$, the subscript ℓ for a component-wise logarithm and $e = (1,\ldots,1)^T \in \mathbb{R}^d$, the first-order optimality condition says

$$t\Sigma^{-1}(x^+-\mu)+(X^+_\ell-\bar{X}_\ell)e = 0 \implies t\Sigma^{-1}(x^+-\mu) = (\bar{X}(X^+)^{-1})_\ell e$$

If Σ is diagonal, the optimization problem is separable and the above reduces to a somewhat closed-form expression in terms of the Lambert W function. We may assume $d = 1$ by Remark 5.1, in which case we obtain

$$x^+ = \frac{\sigma}{t}W\left(\frac{t}{\sigma}\bar{x}e^{\frac{t\mu}{\sigma}}\right)$$

This follows from writing

$$\frac{t}{\sigma}(x^+-\mu)+\log\left(\frac{x^+}{\bar{x}}\right) = 0 \Leftrightarrow \frac{t}{\sigma}x^+e^{\frac{t}{\sigma}x^+} = \frac{t}{\sigma}\bar{x}e^{\frac{t\mu}{\sigma}}$$

Multivariate Normal-inverse Gaussian

The Bregman proximal gradient operator is given by

$$x^+ = \operatorname{argmin}\left\{t\left[\alpha\sqrt{\delta^2+(u-\mu)^T\Sigma^{-1}(u-\mu)}-\langle\beta, u-\mu\rangle-\delta\gamma\right]-\sum_{i=1}^{d} u_i\left(\log\left(\frac{u_i}{\bar{x}_i}\right)-1\right) : u \in \mathbb{R}^d\right\}$$

Denoting $X^+ = \operatorname{diag}(x^+) \succ 0$, the subscript ℓ for a component-wise logarithm and $e = (1,\ldots,1)^T \in \mathbb{R}^d$, the first-order optimality condition says

$$\frac{t\alpha\Sigma^{-1}(x^+-\mu)}{\sqrt{\delta^2+(x^+-\mu)^T\Sigma^{-1}(x^+-\mu)}}-t\beta+(X^+_\ell-\bar{X}_\ell)e = 0$$

where we used $\bar{X} = \operatorname{diag}(\bar{x})$.

In general, to find the proximal operator one needs to solve the above nonlinear system of equations. Nevertheless, the computation can be simplified considerably in the univariate case ($d = 1$), i.e., when we consider the normal-inverse Gaussian distribution. Then, the above equation takes the form

$$(t\alpha/\sigma)(x^+-\mu) = (t\beta-\log(x^+/\bar{x}))\sqrt{\delta^2+(x^+-\mu)^2/\sigma}$$

Gamma

The Gamma Cramer function has parameters $\alpha, \beta > 0$. We may assume separability by Remark 5.1, so we are solving a scalar equation. We seek to find

$$\operatorname{argmin}\left\{t\left(\beta u-\alpha-\alpha\log\left(\frac{\beta u}{\alpha}\right)\right)+u\left(\log\left(\frac{u}{\bar{x}}\right)-1\right) : u \in \mathbb{R}\right\}$$

By Theorem 5.2 we know $x^+ \in \operatorname{int}(\operatorname{dom}\psi)^*_P$ (in particular $x^+ > 0$), and so x^+ satisfies the following first-order optimality condition:

$$0 = t\left(\beta - \frac{\alpha}{x^+}\right) + \log\left(\frac{x^+}{\bar{x}}\right)$$

This can be rewritten as

$$t\beta = \frac{\alpha t}{x^+} + \log\left(\frac{\bar{x}}{x^+}\right)$$

Set $z := 1/x^+$ and exponentiate both sides to obtain:

$$\exp(t\beta) = \exp(\alpha t z)\,\bar{x}z$$

Equivalently, $\alpha t z \exp(\alpha t z) = \alpha t \exp(t\beta)/\bar{x}$, which can be solved with the Lambert W funtion:

$$\alpha t z = W\left(\frac{\alpha t \exp(t\beta)}{\bar{x}}\right) \Leftrightarrow x^+ = \frac{\alpha t}{W\left(\frac{\alpha t \exp(t\beta)}{\bar{x}}\right)}$$

Laplace

Due to Remark 5.1 we may consider without loss of generality the case $d = 1$. Then the Bregman proximal operator is given by

$$x^+ = \operatorname{argmin}\left\{t\psi_R^*(u) + u\left(\log\left(\frac{u}{\bar{x}}\right) - 1\right) : u \in \mathbb{R}\right\}$$

where

$$\psi_R^*(u) := \begin{cases} 0, & u = \mu \\ \sqrt{1 + \left(\frac{u-\mu}{b}\right)^2} - 1 + \log\left(2\left(\frac{u-\mu}{b}\right)^{-2}\left[\sqrt{1 + \left(\frac{u-\mu}{b}\right)^2} - 1\right]\right), & u \neq \mu \end{cases}$$

Denoting $\rho = (u - \mu)/b$ we obtain that $x^+ = \mu + b\rho^+$ where ρ^+ is given by

$$\rho^+ = \operatorname{argmin}\left\{t\psi_R^*(\mu + b\rho) + (\mu + b\rho)\left(\log\left(\frac{\mu + b\rho}{\bar{x}}\right) - 1\right) : \rho \in \mathbb{R}\right\}$$

In view of Corollary 5.1, we know that if $\bar{x} = E_R = \mu$ then $x^+ = \bar{x} = \mu$ and $\rho^+ = 0$. In order to evaluate the value of ρ^+ (and consequently x^+) when $\bar{x} \neq \mu$ we can use Theorem 5.1 to guarantee $x^+ > 0$ and so we may employ first-order optimality conditions. Denoting $\xi(\rho) := (\mu + b\rho)\left(\log\left(\frac{\mu+b\rho}{\bar{x}}\right) - 1\right)$ and using (1.1) we obtain that the optimality conditions are given by

$$(\vartheta(\rho)/t - 1)^2 = \rho^2 + 1$$

where $\vartheta(\rho) := 2t - \rho\xi'(\rho) = 2t - b\rho\log\left(\frac{\mu+b\rho}{\bar{x}}\right)$. Plugging this into the expression above we find:

$$\begin{aligned} \left(1 - \frac{2b\rho}{t}\log\left(\frac{\mu + b\rho}{\bar{x}}\right)\right)^2 &= \rho^2 + 1 \\ \Leftrightarrow \quad 0 = \rho^2 - \frac{b^2\rho^2}{t^2}\log^2\left(\frac{\mu + b\rho}{\bar{x}}\right) &+ \frac{2b\rho}{t}\log\left(\frac{\mu + b\rho}{\bar{x}}\right) \\ \Leftrightarrow \quad 0 = \rho - \frac{b^2\rho}{t^2}\log^2\left(\frac{\mu + b\rho}{\bar{x}}\right) &+ \frac{2b}{t}\log\left(\frac{\mu + b\rho}{\bar{x}}\right) \end{aligned}$$

where the last equivalence follows since $\rho^+ \neq 0$ by assumption.

Poisson

The Poisson Cramer function has parameter $\lambda > 0$. We may assume separability by Remark 5.1, so we are solving a scalar equation. We seek to find

$$\operatorname{argmin}\left\{t\left(u\log\left(\frac{u}{\lambda}\right)-u+\lambda\right)+u\left(\log\left(\frac{u}{\bar{x}}\right)-1\right)\,:\,u\in\mathbb{R}\right\}$$

By Theorem 5.2 we know $x^+ \in \operatorname{int}(\operatorname{dom}\psi)^*_P$ (in particular $x^+ > 0$), and so x^+ satisfies the following first-order optimality condition:

$$0 = t\log\left(\frac{x^+}{\lambda}\right)+\log\left(\frac{x^+}{\bar{x}}\right)$$

Equivalently:

$$0 = \log\left(\frac{(x^+)^{t+1}}{\bar{x}\lambda^t}\right)$$

It follows that

$$x^+ = \left(\bar{x}\lambda^t\right)^{1/(t+1)} = \bar{x}^\tau\lambda^{1-\tau}$$

where $\tau := \frac{t}{t+1}$.

Multinomial

Let $p \in \Delta_{(d)}$ such that $\langle e, p\rangle < 1$. Recall that due to Theorem 5.2 we can assume that $p_i > 0$ for all $i = 1, 2, \ldots, d$. Thus, $p \in \operatorname{int}\Delta_{(d)}$. The Bregman proximal gradient operator (1.4) is given by:

$$x^+ = \operatorname{argmin}\left\{t\left[\sum_{i=1}^d y_i\log\left(\frac{y_i}{np_i}\right)+(n-\langle e,y\rangle)\log\left(\frac{n-\langle e,y\rangle}{n(1-\langle e,p\rangle)}\right)\right]\right.$$
$$\left.+\sum_{i=1}^d u_i\left(\log\left(\frac{u_i}{\bar{x}_i}\right)-1\right)\,:\,u\in n\Delta_{(d)}\right\}$$

We can write the Lagrangian

$$L(u;\lambda) = t\left[\sum_{i=1}^d y_i\log\left(\frac{y_i}{np_i}\right)+(n-\langle e,y\rangle)\log\left(\frac{n-\langle e,y\rangle}{n(1-\langle e,p\rangle)}\right)\right]$$
$$+\sum_{i=1}^d u_i\left(\log\left(\frac{u_i}{\bar{x}_i}\right)-1\right)+\lambda(\langle e,u\rangle-n),$$

where $\lambda \geq 0$ is the Lagrange multiplier. By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi^*_R)$. Thus, writing the KKT conditions, and noting that the complementary slackness condition and the $x^+ \in \operatorname{int}(\operatorname{dom}\psi^*_R)$ yields $\lambda = 0$, we obtain that the unique solution is

$$x^+ \in \operatorname{ri} n\Delta_{(d)} : t\log\left(\frac{x_i^+(1-\sum_{j=1}^d p_j)}{p_i(n-\sum_{j=1}^d x_j^+)}\right)+\log\left(\frac{x_i^+}{\bar{x}_i}\right) = 0$$

The last equation can be written equivalently as

$$x_i^+ = \gamma_i \left(n - \sum_{j=1}^{d} x_j^+\right)^{\frac{t}{t+1}} \qquad \text{where} \qquad \gamma_i = \left[\frac{p_i \bar{x}_i{}^{1/t}}{1 - \sum_{j=1}^{d} p_j}\right]^{\frac{t}{t+1}}$$

Denoting $\rho := \sum_{j=1}^{d} x_j^+$ and summing the above for $i = 1, 2, \ldots, d$ yields

$$\rho = (n - \rho)^{\frac{t}{t+1}} \left(\sum_{i=1}^{d} \gamma_i\right). \tag{1.3}$$

To summarize, $x_i^+ = \gamma_i (n - \rho)^{\frac{t}{t+1}}$ where $\rho \in \mathbb{R}$ is a solution to (1.3).

Negative Multinomial

Recall that due to Theorem 5.2, we can assume $p_i > 0$ for all $i = 1, \ldots, d$. The Bregman proximal gradient operator is given by

$$x^+ = \operatorname{argmin} \left\{ t \left[\sum_{i=1}^{d} u_i \log\left(\frac{u_i}{p_i(x_0 + \sum_{i=1}^{d} u_i)}\right) + x_0 \log\left(\frac{x_0}{x_0 + \sum_{i=1}^{d} u_i}\right)\right] \right.$$
$$\left. + \sum_{i=1}^{d} u_i \left(\log\left(\frac{u_i}{\bar{x}_i}\right) - 1\right) : u \in \mathbb{R}_+^d \right\}$$

By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$. Thus, $x^+ \in \mathbb{R}_{++}^d$ must satisfy the first-order optimality conditions, i.e., for all $i \in \{j \in \{1, 2, \ldots, d\} : p_j \neq 0\}$

$$\log\left(\frac{x_i^+}{\bar{x}_i}\right) + t \log\left(\frac{x_i^+}{p_i(x_0 + \sum_{j=1}^{d} x_j^+)}\right) = 0.$$

Discrete Uniform

Due to Remark 5.1 we will consider without loss of generality the case $d = 1$. Thus in order to evaluate the Bregman proximal gradient operator (1.2) we would like to solve

$$x^+ = \operatorname{argmin} \left\{ t\psi_R^*(u) + u \left(\log\left(\frac{u}{\bar{x}}\right) - 1\right) : u \in \mathbb{R} \right\}$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use Corollary 5.1. Using the properties of Legendre type functions we can write

$$x^+ = \nabla h^* \left(\nabla h(\bar{x}) - t\theta^+\right) = \exp\left(\nabla h(\bar{x}) - t\theta^+ - 1\right) = \bar{x} \exp\left(-t\theta^+\right)$$

where

$$\begin{aligned}
\theta^+ &= t^{-1} \operatorname{iconv}^{h^*}_{t\psi_R(\cdot/t)}(\bar{x}) \\
&= t^{-1} \operatorname{argmin} \{ t\psi_R(\theta/t) + h^*(\bar{x} - \theta) : \theta \in \mathbb{R} \} \\
&= t^{-1} \operatorname{argmin} \{ t \log\left(M_R[\theta/t]\right) + \exp\left(\bar{x} - \theta - 1\right) : \theta \in \mathbb{R} \} \\
&= \operatorname{argmin} \{ t \log\left(M_R[\theta]\right) + \exp\left(\bar{x} - t\theta - 1\right) : \theta \in \mathbb{R} \}
\end{aligned}$$

where we recall that

$$M_R[\theta] = \begin{cases} \frac{\exp((b+1)\theta)-\exp(a\theta)}{n(\exp(\theta)-1)}, & \theta \neq 0 \\ 1, & \theta = 0 \end{cases}$$

In view of Corollary 5.1, we know that if $\bar{x} = E_R = (a+b)/2$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq (a+b)/2$ we can consider the first-order optimality conditions

$$\frac{b\exp((b+1)\theta^+) - a\exp(a\theta^+)}{\exp((b+1)\theta^+) - \exp(a\theta^+)} = \frac{\exp(\theta^+)}{\exp(\theta^+) - 1} + t\exp(\bar{x} - t\theta^+ - 1)$$

Continuous Uniform

Due to Remark 5.1 we will consider without loss of generality the case $d = 1$. Thus in order to evaluate the Bregman proximal gradient operator (1.2) we would like to solve

$$x^+ = \operatorname{argmin}\left\{t\psi_R^*(u) + u\left(\log\left(\frac{u}{\bar{x}}\right) - 1\right) : u \in \mathbb{R}\right\}$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use Corollary 5.1. Using the properties of Legendre type functions we can write

$$x^+ = \nabla h^*\left(\nabla h(\bar{x}) - t\theta^+\right) = \exp\left(\nabla h(\bar{x}) - t\theta^+ - 1\right) = \bar{x}\exp\left(-t\theta^+\right)$$

where

$$\begin{aligned} \theta^+ &= t^{-1}\mathrm{iconv}^{h^*}_{t\psi_R(\cdot/t)}(\bar{x}) \\ &= t^{-1}\operatorname{argmin}\left\{t\psi_R(\theta/t) + h^*(\bar{x} - \theta) : \theta \in \mathbb{R}\right\} \\ &= t^{-1}\operatorname{argmin}\left\{t\log\left(M_R[\theta/t]\right) + \exp\left(\bar{x} - \theta - 1\right) : \theta \in \mathbb{R}\right\} \\ &= \operatorname{argmin}\left\{t\log\left(M_R[\theta]\right) + \exp\left(\bar{x} - t\theta - 1\right) : \theta \in \mathbb{R}\right\} \end{aligned}$$

where we recall that

$$M_R[\theta] = \begin{cases} \frac{\exp(b\theta)-\exp(a\theta)}{\theta(b-a)}, & \theta \neq 0 \\ 1, & \theta = 0 \end{cases}$$

In view of Corollary 5.1, we know that if $\bar{x} = E_R = (a+b)/2$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq (a+b)/2$ we can consider the first-order optimality conditions

$$\frac{b\exp(b\theta^+) - a\exp(a\theta^+)}{\exp(b\theta^+) - \exp(a\theta^+)} = \frac{1}{\theta^+} + t\exp(\bar{x} - t\theta^+ - 1)$$

Logistic

Due to Remark 5.1 we will consider without loss of generality the case $d = 1$. Thus in order to evaluate the Bregman proximal gradient operator (1.2) we would like to solve

$$x^+ = \operatorname{argmin}\left\{t\psi_R^*(u) + u\left(\log\left(\frac{u}{\bar{x}}\right) - 1\right) : u \in \mathbb{R}\right\}$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use Corollary 5.1. Using the properties of Legendre type functions we can write

$$x^+ = \nabla h^* \left(\nabla h(\bar{x}) - t\theta^+\right) = \exp\left(\nabla h(\bar{x}) - t\theta^+ - 1\right) = \bar{x}\exp\left(-t\theta^+\right)$$

where

$$\begin{aligned}
\theta^+ &= t^{-1}\mathrm{iconv}^{h^*}_{t\psi_R(\cdot/t)}(\bar{x}) \\
&= t^{-1}\operatorname{argmin}\left\{t\psi_R(\theta/t) + h^*(\bar{x}-\theta) : \theta \in \mathbb{R}\right\} \\
&= t^{-1}\operatorname{argmin}\left\{t\log\left(M_R[\theta/t]\right) + \exp\left(\bar{x}-\theta-1\right) : \theta \in \mathbb{R}\right\} \\
&= \operatorname{argmin}\left\{t\log\left(M_R[\theta]\right) + \exp\left(\bar{x}-t\theta-1\right) : \theta \in \mathbb{R}\right\} \\
&= \operatorname{argmin}\left\{t\log\left(\exp(\mu\theta)\phi_s(\theta)\right) + \exp\left(\bar{x}-t\theta-1\right) : \theta \in (-1/s, 1/s)\right\} \\
&= \operatorname{argmin}\left\{t\log(\phi_s(\theta)) + t\mu\theta + \exp\left(\bar{x}-t\theta-1\right) : \theta \in (-1/s, 1/s)\right\}
\end{aligned}$$

In view of Corollary 5.1, we know that if $\bar{x} = E_R = \mu$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq \mu$ under similar arguments to those we mentioned when evaluating the Cramér function, we can consider the first-order optimality conditions

$$t\frac{\phi_s'(\theta^+)}{\phi_s(\theta^+)} + t\mu - t\exp\left(\bar{x} - t\theta^+ - 1\right) = 0 \Leftrightarrow \frac{1}{\theta^+} + \frac{\pi s}{\tan(-\pi s\theta^+)} + \mu = \exp\left(\bar{x} - t\theta^+ - 1\right)$$

1.3 Smooth Adaptable Kernel - Burg Entropy (Data Fidelity Reference Distribution - Gamma)

Omitting some constant terms we can write the Bregman proximal gradient operator as:

$$\begin{aligned}
x^+ &= \operatorname{argmin}\left\{t\psi_R^*(u) + D_h(u,\bar{x}) : u \in \mathbb{R}^d\right\} \\
&= \operatorname{argmin}\left\{t\psi_R^*(u) + h(u) - \langle\nabla h(\bar{x}), u\rangle : u \in \mathbb{R}^d\right\},
\end{aligned} \tag{1.4}$$

where $h(x) = -\sum_{i=1}^d \log(x_i)$ and $\bar{z}_i := [\nabla h(\bar{x})]_i = -1/\bar{x}_i$ for all $i = 1, 2, \ldots, d$. Namely, throughout we will be solving

$$x^+ = \operatorname{argmin}\left\{t\psi_R^*(u) - \sum_{i=1}^d (\log(u_i) - u_i/\bar{x}_i) : u \in \mathbb{R}^d\right\}.$$

Multivariate Normal

The Bregman proximal gradient operator (1.4) is given by:

$$x^+ = \operatorname{argmin}\left\{\frac{t}{2}(u-\mu)^T\Sigma^{-1}(u-\mu) - \sum_{i=1}^d \log(u_i) - \langle\bar{z}, u\rangle : u \in \mathbb{R}^d\right\}.$$

Denoting $X^+ := \mathrm{diag}\,(x^+) \succ 0$ and $e = (1,1,\dots,1)^T \in \mathbb{R}^d$ we can write the first-order optimality condition as

$$t\Sigma^{-1}(x^+ - \mu) - (X^+)^{-1}e - \bar{z} = 0 \quad \Rightarrow \quad tX^+\Sigma^{-1}(x^+ - \mu) = e + X^+\bar{z} = (I - X^+\bar{X}^{-1})e,$$

where in the above we used $\bar{z} = \nabla h(\bar{x}) = -1/\bar{x}$ and the notation $\bar{X} := \mathrm{diag}\,(\bar{x})$. If Σ is diagonal the optimization problem defining the Bregman proximal operator is separable and the above reduces to a closed form expression. Without the loss of generality consider the case when $d = 1$ (recall Remark 5.1), then

$$x^+ = \frac{(t/\sigma)\mu - 1/\bar{x} + \sqrt{((t/\sigma)\mu - 1/\bar{x})^2 + 4(t/\sigma)}}{2(t/\sigma)}.$$

Multivariate Normal-inverse Gaussian

The Bregman proximal gradient operator (1.4) is given by:

$$x^+ = \operatorname{argmin}\left\{ t\left[\alpha\sqrt{\delta^2 + (u-\mu)^T\Sigma^{-1}(u-\mu)} - \langle \beta, u - \mu\rangle - \delta\gamma\right] - \sum_{i=1}^{d}\log(u_i) - \langle \bar{z}, u\rangle : u \in \mathbb{R}^d \right\}.$$

Denoting $X^+ := \mathrm{diag}\,(x^+) \succ 0$ and $e = (1,1,\dots,1)^T \in \mathbb{R}^d$ we can write the first-order optimality condition as

$$\frac{t\alpha\Sigma^{-1}(x^+ - \mu)}{\sqrt{\delta^2 + (x^+ - \mu)^T\Sigma^{-1}(x^+ - \mu)}} - t\beta - (X^+)^{-1}e + (\bar{X})^{-1}e = 0, \tag{1.5}$$

where in the above we used $\bar{z} = \nabla h(\bar{x}) = -1/\bar{x}$ and the notation $\bar{X} := \mathrm{diag}\,(\bar{x})$. In general, to find the proximal operator one needs to solve the above general nonlinear system of equations. Nevertheless, there are two important scenarios when the computation can be simplified considerably. The first scenario is the univariate case ($d = 1$), i.e., when we consider the normal-inverse Gaussian distribution. Then, the above equation takes the form

$$(t\alpha/\sigma)(x^+ - \mu)x^+ = \big((t\beta - 1/\bar{x})x^+ + 1\big)\sqrt{\delta^2 + (x^+ - \mu)^2/\sigma}. \tag{1.6}$$

Multidimensional models with a Cramér rate function that admits a separable structure $\psi_R^* = \sum_{i=1}^d \psi_{R_i}^*$ where for each $i = 1,\dots,d$ the reference measure R_i is a univariate normal-inverse Gaussian distribution can be addressed by solving d one-dimensional equations of the form (1.6) (recall Remark 5.1). The second scenario that simplifies (1.5) is the case when $\Sigma = \sigma I$ for some $\sigma > 0$. Denoting

$$\rho := \frac{\alpha t/\sigma}{\sqrt{\delta^2 + \|x^+ - \mu\|^2/\sigma}} \geq 0, \tag{1.7}$$

we can write (1.5) in this case as

$$\rho x^+ - (X^+)^{-1}e = t\beta - (\bar{X})^{-1}e + \rho\mu =: w.$$

Multiplying the above by $X^+ \succ 0$ and observing that $X^+w = Wx^+$ where $W = \mathrm{diag}\,(w)$, we obtain

$$\rho X^+x^+Wx^+ - e = 0 \qquad \Rightarrow \qquad \rho(x_i^+)^2 - w_ix^+ - 1 = 0, \quad i = 1,\dots,d.$$

Since it must hold that $x_i^+ \in \mathbb{R}_{++}$ we can conclude that

$$x_i^+ = \frac{w_i + \sqrt{w_i^2 + 4\rho}}{2\rho},$$

and thus

$$\|x^+ - \mu\|^2 = \sum_{i=1}^{d} \left[\frac{w_i + \sqrt{w_i^2 + 4\rho}}{2\rho} - \mu_i \right]^2.$$

Plugging the above into (1.7) and using the definition of w we obtain that ρ is a positive root of

$$(\rho\delta)^2 + \frac{1}{\sigma} \sum_{i=1}^{d} \left[\frac{1}{2} \left(t\beta_i - \frac{1}{\bar{x}_i} + \sqrt{(t\beta_i - 1/\bar{x}_i + \mu_i\rho)^2 + 4\rho} \right) \right]^2 = (\alpha t/\sigma)^2.$$

Gamma

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, the Bregman proximal gradient operator (1.4) is given by:

$$x^+ = \operatorname{argmin} \left\{ t \left[\beta u - \alpha + \alpha \log \left(\frac{\alpha}{\beta u} \right) \right] - \log(u) - \bar{z}u : u \in \mathbb{R} \right\}.$$

By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom} \psi_R^*)$. Thus, x^+ must satisfy the first-order optimality condition

$$t\beta - \bar{z} - \frac{t\alpha + 1}{x^+} = 0 \quad \Rightarrow \quad x^+ = \frac{t\alpha + 1}{t\beta - \bar{z}} = \frac{\bar{x}(t\alpha + 1)}{\bar{x}t\beta + 1},$$

where in the above we used $\bar{z} = \nabla h(\bar{x}) = -1/\bar{x}$.

Laplace

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, the Bregman proximal gradient operator (1.4) is given by:

$$x^+ = \operatorname{argmin} \left\{ t\psi_R^*(u) - \log(u) - \bar{z}u : u \in \mathbb{R} \right\},$$

where

$$\psi_R^*(u) := \begin{cases} 0, & u = \mu, \\ \sqrt{1 + \left(\frac{u-\mu}{b}\right)^2} - 1 + \log \left(2 \left(\frac{u-\mu}{b}\right)^{-2} \left[\sqrt{1 + \left(\frac{u-\mu}{b}\right)^2} - 1 \right] \right), & u \neq \mu. \end{cases}$$

Denoting $\rho = (u - \mu)/b$, we obtain that $x^+ = \mu + b\rho^+$ where ρ^+ is given by

$$\rho^+ \quad = \operatorname{argmin} \left\{ t\psi_R^*(\mu + b\rho) - \log(\mu + b\rho) - \bar{z}(\mu + b\rho) : \rho \in \mathbb{R} \right\}.$$

In view of Corollary 5.1 we know that if $\bar{x} = E_R = \mu$ then $x^+ = \bar{x} = \mu$ and $\rho^+ = 0$. In order to evaluate the value of ρ^+, and as a result of x^+, when $\bar{x} \neq \mu$ we can consider the first-order optimality conditions. Denoting $\xi(\rho) := -\log(\mu + b\rho) - \bar{z}(\mu + b\rho)$ and using (1.1) we obtain that the optimality conditions are given by

$$(\vartheta(\rho)/t - 1)^2 = \rho^2 + 1,$$

with $\vartheta(\rho) := 2t - \rho\xi'(\rho) = 2t + b\rho/(\mu + b\rho) + \bar{z}b\rho$. Plugging the latter expression into the equation above we obtain that

$$\begin{aligned}
& \left(1 + \frac{b\rho}{t(\mu+b\rho)} + \frac{\bar{z}b\rho}{t}\right)^2 = \rho^2 + 1 \\
\Leftrightarrow\quad & 1 + \left(\frac{b\rho}{t(\mu+b\rho)} + \frac{\bar{z}b\rho}{t}\right)^2 + 2\left(\frac{b\rho}{t(\mu+b\rho)} + \frac{\bar{z}b\rho}{t}\right) = \rho^2 + 1 \\
\Leftrightarrow\quad & \rho\left(\frac{b + \bar{z}b(\mu+b\rho)}{t(\mu+b\rho)}\right)^2 + 2\left(\frac{b + \bar{z}b(\mu+b\rho)}{t(\mu+b\rho)}\right) = \rho \\
\Leftrightarrow\quad & \rho(b + \bar{z}b(\mu+b\rho))^2 + 2t(\mu+b\rho)(b + \bar{z}b(\mu+b\rho)) = \rho t^2(\mu+b\rho)^2 \\
\Leftrightarrow\quad & \rho b^2 + 2\rho\bar{z}b^2(\mu+b\rho) + \rho\bar{z}^2b^2(\mu+b\rho)^2 + 2tb(\mu+b\rho) + 2\bar{z}tb(\mu+b\rho)^2 = \rho t^2(\mu+b\rho)^2 \\
\Leftrightarrow\quad & 0 = \rho^3\left[b^2(\bar{z}^2b^2 - t^2)\right] + \rho^2\left[2b(\bar{z}b^2(1+t) + \mu(\bar{z}^2b^2 - t^2))\right] \\
& \quad + \rho\left[b^2((1+\bar{z}\mu)^2 + 2t(1+2\bar{z}\mu)) + t^2\mu^2\right] - 2tb\mu(1+\bar{z}\mu) \\
\Leftrightarrow\quad & \alpha_1\rho^3 + \alpha_2\rho^2 + \alpha_3\rho + \alpha_4 = 0,
\end{aligned}$$

where $\alpha = [b^2(\bar{z}^2b^2 - t^2), 2b(\bar{z}b^2(t+1) + \mu(\bar{z}^2b^2 - t^2)), b^2((1+\bar{z}\mu)^2 + 2t(1+2\bar{z}\mu)) - t^2\mu^2, 2tb\mu(1+\bar{z}\mu)]^T$.

To summarize,

$$x^+ = \begin{cases} \mu, & \mu = \bar{x}, \\ \mu + b\rho^+, & \mu \neq \bar{x}, \end{cases}$$

where ρ^+ is the unique real root of the cubic equation $\alpha_1\rho^3 + \alpha_2\rho^2 + \alpha_3\rho + \alpha_4 = 0$ with coefficients:

$$\begin{aligned}
\alpha_1 &= b^2((b/\bar{x})^2 - t^2), & \alpha_3 &= b^2((1 - \mu/\bar{x})^2 + 2t(1 - 2\mu/\bar{x})) - t^2\mu^2 \\
\alpha_2 &= 2b(\mu((b/\bar{x})^2 - t^2) - b^2(t+1)/\bar{x}), & \alpha_4 &= 2tb\mu(1 - \mu/\bar{x}),
\end{aligned}$$

where in the above we used $\bar{z} = \nabla h(\bar{x}) = -1/\bar{x}$.

Poisson

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, the Bregman proximal gradient operator (1.4) is given by:

$$x^+ = \operatorname{argmin}\left\{t\left[u\log(u/\lambda) - u + \lambda\right] - \log(u) - \bar{z}u : u \in \mathbb{R}\right\}.$$

By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$. Thus, x^+ must satisfy the first-order optimality condition

$$t\log\left(\frac{u}{\lambda}\right) = \frac{1}{u} - \frac{1}{\bar{x}},$$

where in the above we used $\bar{z} = \nabla h(\bar{x}) = -1/\bar{x}$.

Multinomial

Let $p \in \Delta_{(d)}$ such that $\langle e, p\rangle < 1$. Recall that due to Theorem 5.2 we can assume that $p_i > 0$ for all $i = 1, 2, \ldots, d$. Thus, $p \in \operatorname{int}\Delta_{(d)}$. The Bregman proximal gradient operator (1.4) is given by:

$$\begin{aligned} x^+ = \operatorname{argmin}\Bigg\{ t\left[\sum_{i=1}^{d} y_i \log\left(\frac{y_i}{np_i}\right) + (n - \langle e, y\rangle)\log\left(\frac{n - \langle e, y\rangle}{n(1 - \langle e, p\rangle)}\right)\right] \\ - \sum_{i=1}^{d}\log(u_i) - \langle \bar{z}, u\rangle : u \in n\Delta_{(d)} \Bigg\}. \end{aligned}$$

We can write the Lagrangian

$$\begin{aligned} L(u; \lambda) = t\left[\sum_{i=1}^{d} y_i \log\left(\frac{y_i}{np_i}\right) + (n - \langle e, y\rangle)\log\left(\frac{n - \langle e, y\rangle}{n(1 - \langle e, p\rangle)}\right)\right] \\ - \sum_{i=1}^{d}\log(u_i) - \langle \bar{z}, u\rangle + \lambda(\langle e, u\rangle - n), \end{aligned}$$

where $\lambda \geq 0$ is the Lagrange multiplier. By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$. Thus, writing the KKT conditions, and noting that the complementary slackness condition and the $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$ yields $\lambda = 0$, we obtain that for each $i = 1, 2, \ldots, d$ the unique solution is

$$x^+ \in \operatorname{ri} n\Delta_{(d)} : \quad t\log\left(\frac{x_i^+(1 - \sum_{j=1}^{d} p_j)}{p_i(n - \sum_{j=1}^{d} x_j^+)}\right) = \frac{1}{x_i^+} - \frac{1}{\bar{x}_i}, \qquad i = 1, 2, \ldots, d,$$

where in the above we used $\bar{z}_i = [\nabla h(\bar{x})]_i = -1/\bar{x}_i$.

Negative Multinomial

Recall that due to Theorem 5.2 we can assume that $p_i > 0$ for all $i = 1, 2, \ldots, d$. The Bregman proximal gradient operator (1.4) is given by:

$$\begin{aligned} x^+ &= \operatorname{argmin}\Bigg\{ t\left[\sum_{i=1}^{d} u_i \log\left(\frac{u_i}{p_i(x_0 + \sum_{i=1}^{d} u_i)}\right) + x_0 \log\left(\frac{x_0}{x_0 + \sum_{i=1}^{d} u_i}\right)\right] \\ &\qquad - \sum_{i=1}^{d}\log(u_i) - \langle \bar{z}, u\rangle : u \in \mathbb{R}_+^d \Bigg\}. \end{aligned}$$

By Theorem 5.2 we know that $x^+ \in \operatorname{int}(\operatorname{dom}\psi_R^*)$. Thus, $x^+ \in \mathbb{R}_{++}^d$ must satisfy the first-order optimality conditions, i.e., for all $i \in \{j \in \{1, 2, \ldots, d\} : p_j \neq 0\}$

$$t\log\left(\frac{x_i^+}{p_i(x_0 + \sum_{i=1}^{d} x_i^+)}\right) = \frac{1}{x_i^+} - \frac{1}{\bar{x}_i},$$

where in the above we used $\bar{z}_i = [\nabla h(\bar{x})]_i = -1/\bar{x}_i$.

Discrete Uniform

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, in order to evaluate the Bregman proximal gradient operator (1.4) we would like to solve

$$x^+ \quad = \operatorname{argmin} \left\{ t\psi_R^*(u) - \log(u) - \bar{z}u : u \in \mathbb{R} \right\}.$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use Corollary 5.1. Hence, recalling that $\bar{z} = \nabla h(\bar{x}) = -1/\bar{x}$ we can write

$$x^+ = \nabla h^* \left(\bar{z} - t\theta^+ \right) = \frac{1}{t\theta^+ - \bar{z}} = \frac{\bar{x}}{\bar{x}t\theta^+ + 1},$$

with

$$\begin{aligned} \theta^+ = t^{-1}\mathrm{iconv}_{t\psi_R(\cdot/t)}^{h^*}(\bar{x}) \quad &= t^{-1} \operatorname{argmin} \left\{ t\psi_R(\cdot/t) + h^*(\bar{z} - \theta) : \theta \in \mathbb{R} \right\} \\ &= t^{-1} \operatorname{argmin} \left\{ t \log \left(M_R[\theta/t] \right) - 1 - \log(\theta - \bar{z}) : \theta \in \mathbb{R} \right\} \\ &= \operatorname{argmin} \left\{ t \log \left(M_R[\theta] \right) - 1 - \log(t\theta - \bar{z}) : \theta \in \mathbb{R} \right\}, \end{aligned}$$

where we recall that

$$M_R[\theta] = \begin{cases} \frac{\exp((b+1)\theta) - \exp(a\theta)}{n(\exp(\theta)-1)}, & \theta \neq 0, \\ 1, & \theta = 0. \end{cases}$$

In view of Corollary 5.1 we know that if $\bar{x} = E_R = (a+b)/2$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq (a+b)/2$ we can consider the first-order optimality conditions

$$\frac{(b+1)\exp\left((b+1)\theta\right) - a\exp\left(a\theta\right)}{(\exp\left((b+1)\theta\right) - \exp\left(a\theta\right)} = \frac{\exp\left(\theta\right)}{\exp\left(\theta\right) - 1} + \frac{1}{t\theta - \bar{z}} = \frac{\exp\left(\theta\right)}{\exp\left(\theta\right) - 1} + \frac{\bar{x}}{\bar{x}t\theta^+ + 1}.$$

Continuous Uniform

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, in order to evaluate the Bregman proximal gradient operator (1.4) we would like to solve

$$x^+ \quad = \operatorname{argmin} \left\{ t\psi_R^*(u) - \log(u) - \bar{z}u : u \in \mathbb{R} \right\}.$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use Corollary 5.1. Hence, recalling that $\bar{z} = \nabla h(\bar{x}) = -1/\bar{x}$ we can write

$$x^+ = \nabla h^* \left(\bar{z} - t\theta^+ \right) = \frac{1}{t\theta^+ - \bar{z}} = \frac{\bar{x}}{t\bar{x}\theta^+ + 1},$$

with

$$\begin{aligned} \theta^+ = t^{-1}\mathrm{iconv}_{t\psi_R(\cdot/t)}^{h^*}(\bar{x}) \quad &= t^{-1} \operatorname{argmin} \left\{ t\psi_R(\cdot/t) + h^*(\bar{z} - \theta) : \theta \in \mathbb{R} \right\} \\ &= t^{-1} \operatorname{argmin} \left\{ t \log \left(M_R[\theta/t] \right) - 1 - \log(\theta - \bar{z}) : \theta \in \mathbb{R} \right\} \\ &= \operatorname{argmin} \left\{ t \log \left(M_R[\theta] \right) - 1 - \log(t\theta - \bar{z}) : \theta \in \mathbb{R} \right\}, \end{aligned}$$

where we recall that

$$M_R[\theta] = \begin{cases} \frac{\exp(b\theta)-\exp(a\theta)}{\theta(b-a)}, & \theta \neq 0, \\ 1, & \theta = 0. \end{cases}$$

In view of Corollary 5.1 we know that if $\bar{x} = E_R = (a+b)/2$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq (a+b)/2$ we can consider the first-order optimality conditions

$$\frac{b\exp(b\theta^+) - a\exp(a\theta^+)}{\exp(b\theta^+) - \exp(a\theta^+)} = \frac{1}{\theta^+} + \frac{1}{t\theta^+ - \bar{z}} = \frac{1}{\theta^+} + \frac{\bar{x}}{\bar{x}t\theta^+ + 1}.$$

Logistic

Due to Remark 5.1 we will consider, without the loss of generality, the case when $d = 1$. Thus, in order to evaluate the Bregman proximal gradient operator (1.4) we would like to solve

$$x^+ \quad = \operatorname{argmin}\left\{t\psi_R^*(u) - \log(u) - \bar{z}u : u \in \mathbb{R}\right\}.$$

Since the Cramér rate function ψ_R^* does not admit a closed form expression, we will use Corollary 5.1. Hence, recalling that $\bar{z} = \nabla h(\bar{x}) = -1/\bar{x}$ we can write

$$x^+ = \nabla h^*\left(\bar{z} - t\theta^+\right) = \frac{1}{t\theta^+ - \bar{z}} = \frac{\bar{x}}{\bar{x}t\theta^+ + 1},$$

with

$$\begin{aligned} \theta^+ = t^{-1}\mathrm{iconv}^{h^*}_{t\psi_R(\cdot/t)}(\bar{x}) \quad &= t^{-1}\operatorname{argmin}\left\{t\psi_R(\cdot/t) + h^*(\bar{z} - \theta) : \theta \in \mathbb{R}\right\} \\ &= t^{-1}\operatorname{argmin}\left\{t\log\left(M_R[\theta/t]\right) - 1 - \log(\theta - \bar{z}) : \theta \in \mathbb{R}\right\} \\ &= \operatorname{argmin}\left\{t\log\left(M_R[\theta]\right) - 1 - \log(t\theta - \bar{z}) : \theta \in \mathbb{R}\right\} \\ &= \operatorname{argmin}\left\{t\log\left(\exp(\mu\theta)\phi_s(\theta)\right) - 1 - \log(t\theta - \bar{z}) : \theta \in (-1/s, 1/s)\right\} \\ &= \operatorname{argmin}\left\{t\log\left(\phi_s(\theta)\right) + t\mu\theta - 1 - \log(t\theta - \bar{z}) : \theta \in (-1/s, 1/s)\right\}. \end{aligned}$$

In view of Corollary 5.1 we know that if $\bar{x} = E_R = \mu$ then $\theta^+ = 0$. In order to evaluate the value of θ^+ when $\bar{x} \neq \mu$, under similar arguments to those we mentioned when evaluated the Cramér rate function, we can write the first order optimality conditions of the above problem

$$\frac{t\phi_s'(\theta^+)}{\phi_s(\theta^+)} + t\mu = \frac{t}{t\theta^+ - \bar{z}} \qquad \Rightarrow \qquad \frac{1}{\theta^+} + \frac{\pi s}{\tan\left(-\pi s\theta^+\right)} + \mu = \frac{1}{t\theta^+ - \bar{z}} = \frac{\bar{x}}{\bar{x}t\theta^+ + 1}.$$

References

[1] A. Beck. First-order methods in optimization. *SIAM*, 25, (2017).