Cross Entropy under Linear Constraints

Nicholas Barnfield   Michael P. Friedlander   Tim Hoheisel

2024-04-03

Problem statement

Define the cross-entropy and log-exponential functions by

$$
\begin{aligned}
\operatorname{cent}\ (p \mid q) &= \sum_{i=1}^{n} p_i \log(p_i / q_i), \\
\operatorname{logexp}\ (p \mid q) &= \log \sum_{i=1}^{n} q_i \exp(p_i),
\end{aligned}
$$

which form a conjugate pair. We consider the pair of dual optimization problems

$$
\begin{aligned}
&\min_{p,\,y} \left\{ \operatorname{cent}\ (p \mid q) + \lambda/2 \,\| y \|^2 \mid Ax + \lambda y = b \right\} \\
&\min_{y} \left\{ \operatorname{logexp}\ (A^T y \mid q) + \lambda/2 \,\| y \|^2 - \langle b, y \rangle \right\},
\end{aligned}
$$

where A is an m-by-n matrix, b is an m-vector, and λ is a nonnegative regularization parameter.

Implicit approach

Optimality conditions

A pair (p, y) is primal-dual optimal if and only if it satisfies the optimality conditions

$$
\begin{aligned}
b &= Ax + \lambda y, \\
x_j &= \frac{1}{\sigma(x)} q_j \exp\big(\langle a_j, y \rangle\big), \quad j \in [n], \\
\sigma(x) &= \sum_{j=1}^{n} q_j \exp\big(\langle a_j, y \rangle\big),
\end{aligned}
$$

Testing

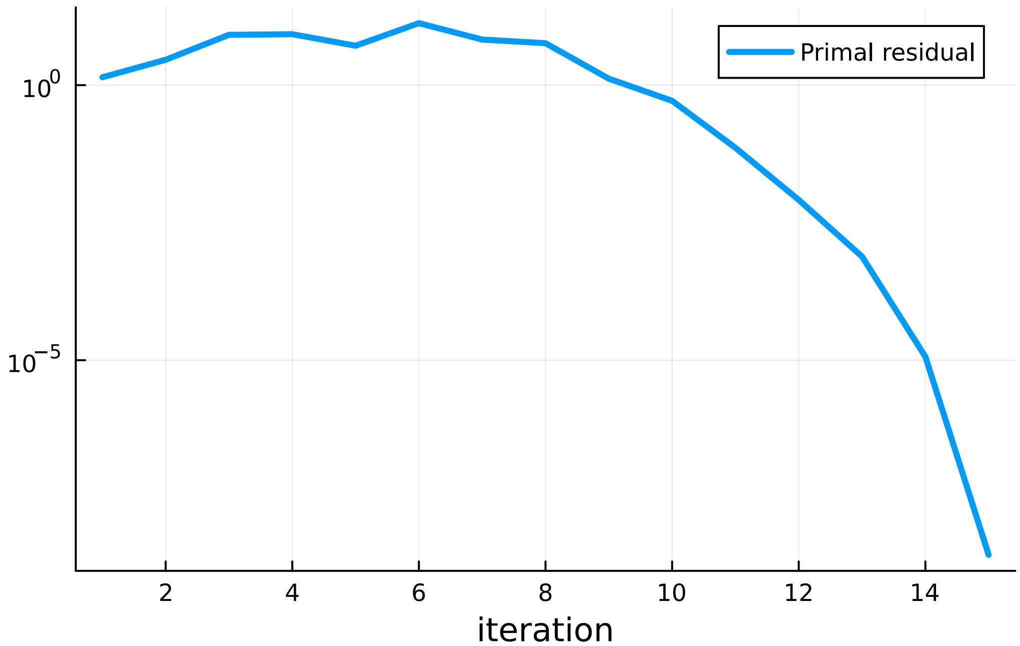
```julia
using Plots, KLLS, Random
```

```julia
Random.seed!(1234)
m, n = 200, 300
q = fill(1/n, n)
A = randn(m, n)
b = A*q + 0.1*randn(m)
λ = 1e-2;
```

```julia
p, y, trace = newton_opt(A, b, q, λ);
```

```
iter:          obj       |grd|  line its
   0: -3.5397e-15  1.3900e+00        0
   1: -2.6795e+00  2.8947e+00        1
   2: -7.7034e+00  8.2307e+00        2
   3: -1.0797e+01  8.4474e+00        4
   4: -1.2758e+01  5.2122e+00        5
   5: -1.2860e+01  1.3412e+01        1
   6: -1.7295e+01  6.7487e+00        7
   7: -1.7892e+01  5.7937e+00        4
   8: -1.8956e+01  1.3129e+00        2
   9: -1.9204e+01  5.1984e-01        1
  10: -1.9270e+01  7.2606e-02        1
  11: -1.9272e+01  8.2058e-03        1
  12: -1.9272e+01  7.6512e-04        1
  13: -1.9272e+01  1.1550e-05        1
  14: -1.9272e+01  2.8837e-09        1
```

```julia
plot(trace.pResid, yscale=:log10, label="Primal residual", xlabel="iteration",
lw=3)
```



Primal residual
10^{0}
10^{-5}
2
4
6
8
10
12
14
iteration