MATH 6397

Rosalind Finley, Alexis Tanner

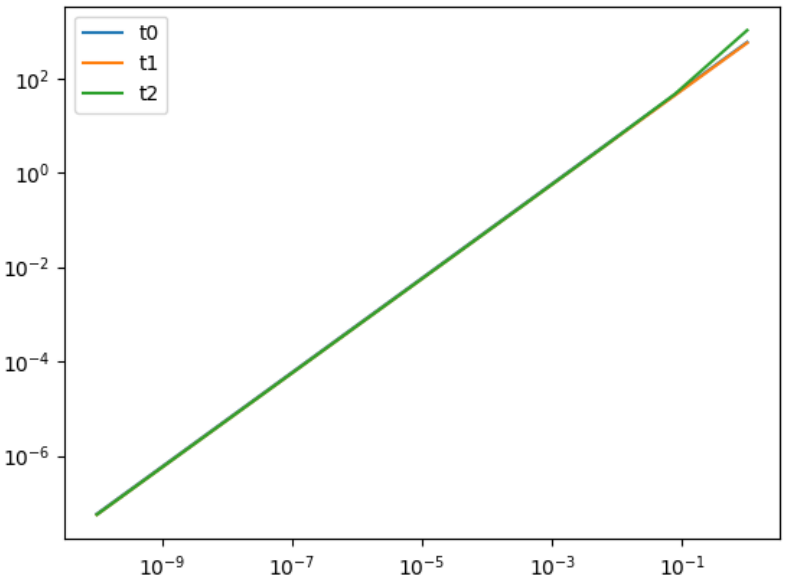
Problem Set 1

1. We are given $f(x) = \frac{1}{2}x^T Q x + b^T x + c$.

   Therefore,

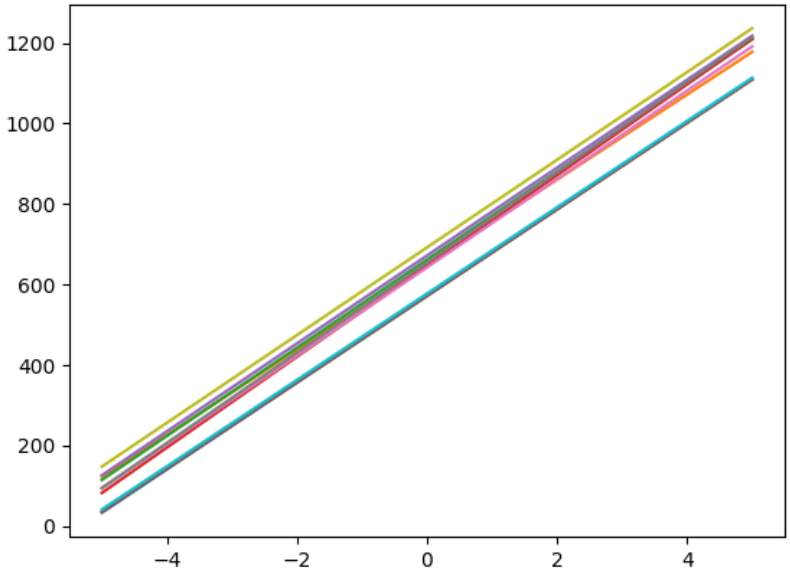
$$\nabla f(x) = \frac{1}{2}(Q^T + Q)x + b,$$
$$\nabla^2 f(x) = \frac{1}{2}(Q^T + Q).$$

   a.) Derivative Check:



   Comments: We derive the gradient and Hessian by hand in 1a_gradient_Hessian.pdf. The quadobj_deriv_check.py output confirms our results.

b.) Convexity Check:



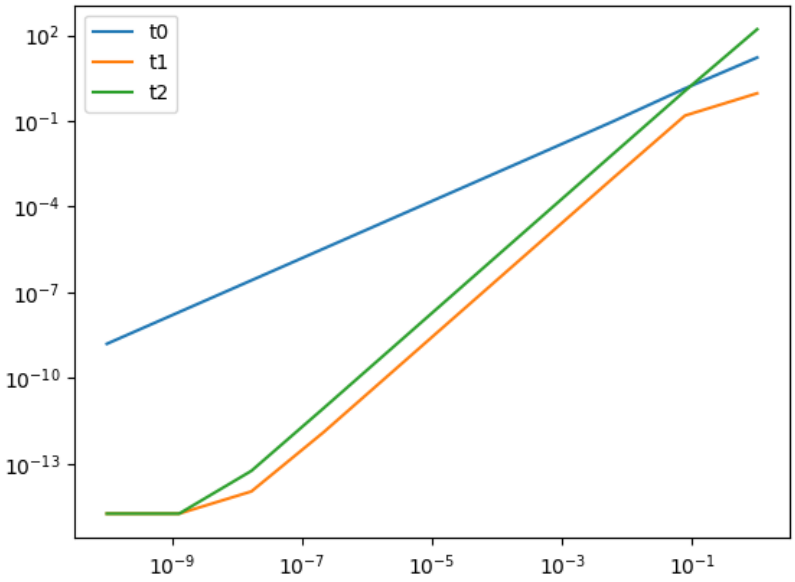
Comments: The quadobj_cvx_check.py output confirms that f is a convex function.

2. We are given: $\underset{x\in\mathbb{R}^n}{\text{minimize}} f(x)$, where $f(x) = \{\frac{1}{2}\|sin(Ax) - b\|_2^2 + \frac{\beta}{2}\|Lx\|_2^2\}$.

We derive the gradient and Hessian:

$$\nabla f(x) = A^T diag(cos(Ax))(sin(Ax) - b) + \beta L^T Lx,$$
$$\nabla^2 f(x) = A^T A\{1 - 2sin(Ax)(sin(Ax))^T + b^T sin(Ax)\} + \beta L^T L.$$

Derivative Check:

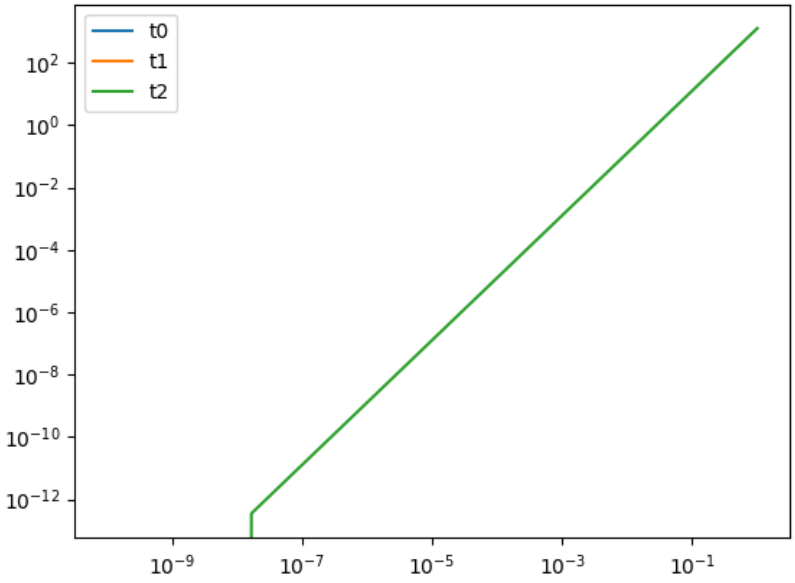


3. We are given the optimization problem $\underset{x\in\mathbb{R}^n}{\text{minimize}} f(x)$, where $f(x) = \{\frac{1}{2}\|\sigma(YX) - C\|_F^2\}$.

a.) We derive the partial derivative, whereby we obtain the gradient:

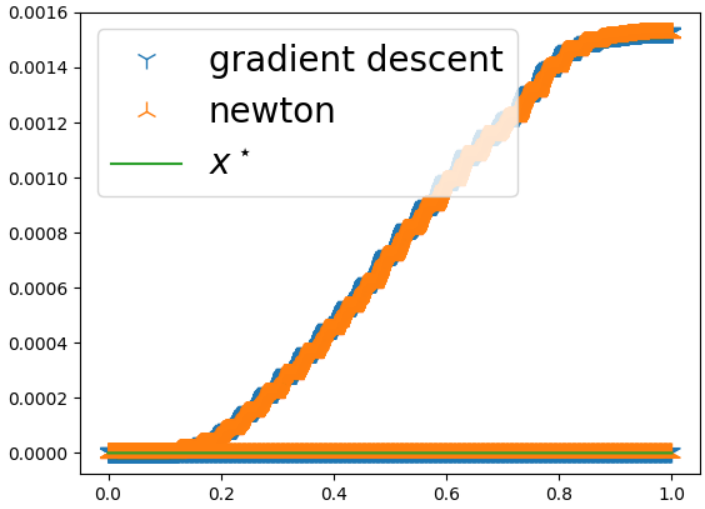
$$\frac{\partial f}{\partial x_{i,j}} = \sum_{k=1}^{m} y_{ki}(1 - \sigma^2(\sum_{l=1}^{n} y_{kl}x_{lj}))(\sigma(\sum_{l=1}^{n} y_{kl}x_{lj}) - c_{kj}),$$
$$\nabla f(x) = Y^T((1 - tanh^2(YX)) \odot (tanh(YX) - C).$$

Derivative Check:




```
-------------------------------------------------
executing derivative check
-------------------------------------------------
h           t1          t2          t3
-------------------------------------------------
1.000000e+00 0.000000e+00 0.000000e+00 1.291424e+03
7.742637e-02 0.000000e+00 0.000000e+00 7.741884e+00
5.994843e-03 0.000000e+00 0.000000e+00 4.641138e-02
4.641589e-04 0.000000e+00 0.000000e+00 2.782289e-04
3.593814e-05 0.000000e+00 0.000000e+00 1.667938e-06
2.782559e-06 0.000000e+00 0.000000e+00 9.999027e-09
2.154435e-07 0.000000e+00 0.000000e+00 5.994139e-11
1.668101e-08 0.000000e+00 0.000000e+00 3.623768e-13
1.291550e-09 0.000000e+00 0.000000e+00 0.000000e+00
1.000000e-10 0.000000e+00 0.000000e+00 0.000000e+00
-------------------------------------------------
```

b.) Gradient Descent



Training classification accuracy (1 iteration):

```
executing gradient descent
------------------------------------------------------
  iter          ||df||      ||df||_rel            step
------------------------------------------------------
     0    4.891082e+02    2.268506e-01    1.220703e-04
     1    2.877898e+02    1.334783e-01    2.441406e-04
>> maximum number of iterations ( 1 ) reached
------------------------------------------------------
prediction accuracy:  8.0 ( 8.000000e-02 )
```

Training classification accuracy (100 iterations):

```
>> maximum number of iterations ( 100 ) reached
------------------------------------------------------
prediction accuracy:  16.0 ( 1.600000e-01 )
```

Test classification accuracy (1 iteration):

```
------------------------------------------------------
executing gradient descent
------------------------------------------------------
  iter          ||df||      ||df||_rel            step
------------------------------------------------------
     0    1.065512e-11    4.777203e-15    1.953125e-03
>> solver converged: 1.065512e-11 < 2.230411e-03
------------------------------------------------------
prediction accuracy:  14.000000000000002 ( 1.400000e-01 )
```

Test classification accuracy (100 iterations):

```
>> maximum number of iterations ( 100 ) reached
------------------------------------------------------
prediction accuracy:  8.0 ( 8.000000e-02 )
```