Finals

Denis Zuenko

October 28, 2018

1 Problem 1

1.1 (a)

$$\min_x \|x = 2\|$$
$$\text{subject to } \|x\|_* \leq 1, x \in \mathbb{R}^n$$

It is easy to see that the point of intersection lies on the affine set, where: $x_1 = x_2 = \cdots = x_n$. In that case minimum value corresponds to strict inequality constriction $\|x\|_q = 1$.
Let all components of x be equal to each other, then:

$$\|x\|_q = 1 \implies x_i = \frac{1}{n^{1/q}}$$

Hence dual problem:

$$L(x, \lambda) = \|x - 2\|_p + \lambda(\|x\|_* - 1)$$
$$g(\lambda) = infxL(x, \lambda)$$
$$\lambda \geq 0$$

Strong duality holds as points with $\|x\| < 1$ - are feasible.

1.2 (b)

$$\min_x \sum_{i=1}^N x_i log(\frac{x_i}{y_i})$$
$$\text{subject to } Ax = b$$
$$1^T x = 1$$
$$x \prec 0$$

Lagrangian:

$$L(x, \lambda, v) = \sum_{i=1}^N x_i log(\frac{x_i}{y_i}) - \lambda^T x + v_1(1^T - 1) + v_2(Ax - b)$$

where $g(\lambda, v_1, v_2) = infxL$

Therefore:

$$\frac{dL}{dx_i} = log(\frac{x_i}{y_i} - 1 - \lambda_i v_{1,i} + v_{2,i}^T A^i = 0 \implies x_i^* = exp^{\lambda_i + 1 + v_i + v_2^T A_i + log(y_i)}$$

Dual problem:

$$g(\lambda, v_1, v_2) = \sum_{i=1}^N x_i^* log(\frac{x_i^*}{y_i} - \lambda^T x^* + v_1(1^T - 1) + v_2(Ax * -b), \text{ where } \lambda \geq 0$$

Strong duality, Slater's condition.

1.3 (c)

$$\min_x \sum_{i=1}^N \Phi(x_i), \qquad \Phi(x_i) = \frac{|x_i|}{c - |x_i|}$$

subject to $Ax = b$

$x - c_1 \prec 0$

Lagrangian:

$L(x, \lambda, v) = \sum_{i=1}^{N} \Phi + \sum_{i=1}^{N} \lambda_{1,i}(x_i - c) - \sum_{i=1}^{N}(\lambda_{2,i}(x_i - c) + v^T(Ax - b)$

where $\frac{|x_i|}{c-|x_i|} = \frac{\sqrt{x_i^2}}{c-\sqrt{x_i^2}}$

After differentiation we get two solutions:

$x_i = -\sqrt{\frac{c}{\lambda_{1,i}-\lambda_{2,i}+v^T A_i}} - c, \quad c^{-1} < \lambda_{1,i} - \lambda_{2,i} + v^T A_i < 0$

$x_i = -\sqrt{\frac{-c}{\lambda_{1,i}-\lambda_{2,i}+v^T A_i}} + c, \quad c^{-1} < \lambda_{1,i} - \lambda_{2,i} + v^T A_i < \frac{-1}{c}$

They exist only if c ¡ 0, hence there is no solution in 0.

2 Problem 2

2.1 (a)

1. Let's consider a function $g(y_1, y_2) = x_1 y_1 + \sqrt{\gamma} x_2 y_2$ and find its supremum w.r.t y_1 and y_2, given two constraints: $y_1^2 + y_2^2 \leq 1$ and $y_1 \geq \frac{1}{\sqrt{1+\gamma}}$, by computing Lagrangian:

$$L(y_1, y_2, \lambda, \mu) = x_1 y_1 + \sqrt{\gamma} x_2 y_2 - \lambda(y_1^2 + y_2^2 - 1) - \mu(1/\sqrt{1+\gamma} - y_1)$$

Let's write KKT conditions for finding exrremum cases.

$$\frac{\partial L}{\partial y_1} = x_1 - 2y_1\lambda + \mu = 0 \tag{1}$$

$$\frac{\partial L}{\partial y_2} = \gamma x_2 - 2y_2\lambda = 0 \tag{2}$$

$$\lambda(y_1^2 + y_2^2 - 1) = 0 \tag{3}$$

$$\mu(1/\sqrt{1+\gamma} - y_1) = 0 \tag{4}$$

$$\mu \geq 0, \quad \lambda \geq 0, \quad y_1 \geq \frac{1}{\sqrt{1+\gamma}}, \quad y_1^2 + y_2^2 \leq 1$$

Let's consider 4 possible cases

2. $\lambda = 0, \qquad \mu = 0, \qquad y_1 > \frac{1}{\sqrt{1+\gamma}}, \qquad y_1^2 + y_2^2 < 1$
Then $x_1 = 0$ and $x_2 = 0$, which leads to $g(y_1, y_2) = 0$, and $\sup_{y_1,y_2} g(y_1, y_2) = 0$

3. $\lambda = 0, \qquad \mu > 0, \qquad y_1 = \frac{1}{\sqrt{1+\gamma}}, \qquad y_1^2 + y_2^2 < 1$
Then $x_2 = 0$, $x_1 = -\mu$, $y_1 = \frac{1}{\sqrt{1+\gamma}}$, and $\sup_{y_1,y_2} g(y_1, y_2) = \sup_{y_1,y_2} \frac{x_1}{\sqrt{1+\gamma}} = \frac{x_1}{\sqrt{1+\gamma}}$

4. $\lambda > 0, \qquad \mu = 0, \qquad y_1 < \frac{1}{\sqrt{1+\gamma}}, \qquad y_1^2 + y_2^2 = 1$
Then from (2.1), we have $y_1 = \frac{x_1}{2\lambda}$ and from (2.2), we have $y_2 = \frac{\sqrt{\gamma}x_2}{2\lambda}$.
Using $y_1^2 + y_2^2 = 1$, we find $\lambda = \frac{\sqrt{x_1^2+\gamma x_2^2}}{2}$, $y_1 = \frac{x_1}{\sqrt{x_1^2+\gamma x_2^2}}$, $y_2 = \frac{\gamma x_2}{\sqrt{x_1^2+\gamma x_2^2}}$

$$\sup_{y_1,y_2} g(y_1, y_2) = x_1^2 + \gamma x_2^2$$

This is only possible if $y_1 < \frac{1}{\sqrt{1+\gamma}}$, or $\frac{x_1}{\sqrt{x_1^2+\gamma x_2^2}} < \frac{1}{\sqrt{1+\gamma}}$, or $x_1 > |x_2|$

5. $\lambda > 0, \qquad \mu > 0, \qquad y_1 = \frac{1}{\sqrt{1+\gamma}}, \qquad y_1^2 + y_2^2 = 1$
   Then $y_2^2 = \frac{\gamma}{1+\gamma}$, and using (2.2) and $\lambda > 0$, we have $\lambda = \frac{|x_2|\sqrt{1+\gamma}}{2}$. Using (2.1), we have $\mu = |x_2| - x_1$

$$\sup_{y_1,y_2} g(y_1, y_2) = \sup_{y_1,y_2} \frac{x_1 + \sqrt{\gamma}x_2y_2}{\sqrt{1+\gamma}}$$

   If $x_2 < 0$, or $x_2 = -|x_2|$, then $y_2 = \sqrt{\frac{\gamma}{1+\gamma}}$, and $\sup_{y_1,y_2} g(y_1, y_2) = \frac{x_1+\gamma x_2}{\sqrt{1+\gamma}}$
   If $x_2 > 0$, or $x_2 = |x_2|$, then $y_2 = -\sqrt{\frac{\gamma}{1+\gamma}}$, and $\sup_{y_1,y_2} g(y_1, y_2) = \frac{x_1-\gamma x_2}{\sqrt{1+\gamma}}$
   In any cases $\sup_{y_1,y_2} g(y_1, y_2) = \frac{x_1-\gamma|x_2|}{\sqrt{1+\gamma}}$
   This is possible when $\mu > 0$ or $|x_2| > x_1$

   All cases described above can be concluded as:

$$f(x_1, x_2) = \sup_{y_1,y_2} \left\{ x_1y_1 + \sqrt{\gamma}x_2y_2 \;:\; y_1^2 + y_2^2 \leq 1, y_1 \geq \frac{1}{\sqrt{1+\gamma}} \right\} = \begin{cases} \sqrt{x_1^2 + \gamma x_2^2} & |x_2| < x_1 \\ \frac{x_1-\gamma|x_2|}{\sqrt{1+\gamma}} & otherwise \end{cases}$$

   The definition using sup operator is convex because it is a sum of two linear convex functions and domain is convex. In (y_1, y_2)-space domain is bounded by a circle of radius 1 and vertical line $y_1 = \frac{1}{\sqrt{1+\gamma}}$. Due to $\gamma > 1$, $\frac{1}{\sqrt{1+\gamma}} < 1$, which means the domain is non-zero and it is a segment of the circle.

6. For gradient descent with exact line search we first need to compute partial derivatives of the function with respect to each variable. Since $x^{(k)} = (\gamma, 1)$, which means $x_1^{(k)} > |x_2^{(k)}|$, we use $f(x_1, x_2) = \sqrt{x_1^2 + \gamma x_2^2}$, thus:

$$\frac{\partial f}{\partial x_1} = \frac{x_1}{\sqrt{x_1^2 + \gamma x_2^2}}$$

$$\frac{\partial f}{\partial x_2} = \frac{\gamma x_2}{\sqrt{x_1^2 + \gamma x_2^2}}$$

   According to exact line search, the step size $\eta^{(k)}$ is chosen to minimize f along the ray $x^{(k)} - \eta\nabla f(x^{(k-1)})$. Thus:

$$\eta^{(1)} = \arg\min_{\eta} f(x^{(0)} - \eta\nabla f(x^{(0)})) = \arg\min_{\eta} \sqrt{\left(\gamma - \eta\frac{\gamma}{\sqrt{\gamma^2+1}}\right)^2 + \gamma\left(1 - \eta\frac{\gamma}{\sqrt{\gamma^2+1}}\right)^2}$$

   To find its minimum we zero the derivative of its square w.r.t. $\eta^{(1)}$:

$$2\left(\gamma - \eta^{(1)}\frac{\gamma}{\sqrt{\gamma^2+1}}\right)\frac{\gamma}{\sqrt{\gamma^2+1}} + 2\gamma\left(1 - \eta^{(1)}\frac{\gamma}{\sqrt{\gamma^2+1}}\right)\frac{\gamma}{\sqrt{\gamma^2+1}} = 0$$

   Solving this equation we get $\eta^{(1)} = \frac{2\sqrt{\gamma^2+1}}{\gamma+1}$. Then

$$x_1^{(1)} = x_1^{(0)} - \eta^{(1)}\nabla f(x_1^{(0)}, x_0^{(0)}) = \gamma\frac{\gamma-1}{\gamma+1}$$

$$x_2^{(1)} = x_2^{(0)} - \eta^{(1)}\nabla f(x_1^{(0)}, x_2^{(0)}) = -\frac{\gamma-1}{\gamma+1}$$

As we see $x_1^{(1)} > |x_2^{(1)}|$, for next iteration we also use $f(x_1, x_2) = \sqrt{x_1^2 + \gamma x_2^2}$:

$$\eta^{(2)} = \arg\min_{\eta} f(x^{(1)} - \eta \nabla f(x^{(1)})) = \arg\min_{\eta} \sqrt{\left(\gamma \frac{\gamma-1}{\gamma+1} - \eta \frac{\gamma}{\sqrt{\gamma^2+\gamma}}\right)^2 + \gamma \left(-\left(\frac{\gamma-1}{\gamma+1}\right) + \eta \frac{\gamma}{\sqrt{\gamma^2 +}}\right.}$$

$$2\left(\gamma \frac{\gamma-1}{\gamma+1} - \eta^{(2)} \frac{\gamma}{\sqrt{\gamma^2+\gamma}}\right) \frac{\gamma}{\sqrt{\gamma^2+\gamma}} - 2\gamma \left(-\left(\frac{\gamma-1}{\gamma+1}\right) + \eta^{(2)} \frac{\gamma}{\sqrt{\gamma^2+\gamma}}\right) \frac{\gamma}{\sqrt{\gamma^2+\gamma}} = 0$$

Solving this equation we get $\eta^{(2)} = \frac{2\gamma(\gamma-1)}{(\gamma+1)\sqrt{\gamma^2+\gamma}}$. Then:

$$x_1^{(2)} = x_1^{(1)} - \eta^{(2)} \nabla f(x_1^{(1)}, x_0^{(1)}) = \gamma \left(\frac{\gamma-1}{\gamma+1}\right)^2$$

$$x_2^{(2)} = x_2^{(1)} - \eta^{(2)} \nabla f(x_1^{(1)}, x_2^{(1)}) = \left(-\frac{\gamma-1}{\gamma+1}\right)^2$$

It's logical to assume that $x_1^{(k)} = \gamma \left(\frac{\gamma-1}{\gamma+1}\right)^k$,$x_2^{(k)} = \left(-\frac{\gamma-1}{\gamma+1}\right)^k$. Due to $\gamma > 1$ for this assumption $x_1^{(k)} > |x_2^{(k)}|$ and $f(x_1, x_2) = \sqrt{x_1^2 + \gamma x_2^2}$. Let's prove the assumption by induction:

$$\eta^{(k+1)} = \arg\min_{\eta} f(x^{(k)} - \eta \nabla f(x^{(k)})) = \arg\min_{\eta} \sqrt{\left(\gamma \left(\frac{\gamma-1}{\gamma+1}\right)^k - \eta \frac{\gamma}{\sqrt{\gamma^2+\gamma}}\right)^2 + \gamma \left(\left(-\frac{\gamma-1}{\gamma+1}\right)^k +\right.}$$

$$2\left(\gamma \left(\frac{\gamma-1}{\gamma+1}\right)^k - \eta^{(k+1)} \frac{\gamma}{\sqrt{\gamma^2+\gamma}}\right) \frac{\gamma}{\sqrt{\gamma^2+\gamma}} - 2\gamma \left((-1)^k \left(\frac{\gamma-1}{\gamma+1}\right)^k - \eta^{(k+1)} \frac{(-1)^k \gamma}{\sqrt{\gamma^2+\gamma}}\right) \frac{\gamma}{\sqrt{\gamma^2+\gamma}} =$$

Solving this equation we get $\eta^{(k+1)} = \frac{2\sqrt{\gamma^2+1}}{\gamma+1} \left(\frac{\gamma-1}{\gamma+1}\right)^k$. Then:

$$x_1^{(k+1)} = x_1^{(k)} - \eta^{(k+1)} \nabla f(x_1^{(k)}, x_0^{(k)}) = \gamma \left(\frac{\gamma-1}{\gamma+1}\right)^{k+1}$$

$$x_2^{(k+1)} = x_2^{(k)} - \eta^{(k+1)} \nabla f(x_1^{(k)}, x_2^{(k)}) = \left(-\frac{\gamma-1}{\gamma+1}\right)^{k+1}$$

which proves the assumption.
Due to $\frac{\gamma-1}{\gamma+1} < 1$ he limits are:

$$\lim_{k\to\infty} x_1^{(k)} = \lim_{k\to\infty} \gamma \left(\frac{\gamma-1}{\gamma+1}\right)^k = 0$$

$$\lim_{k\to\infty} x_2^{(k)} = \lim_{k\to\infty} \left(-\frac{\gamma-1}{\gamma+1}\right)^k = 0$$

Which means solution converges to $(0, 0)$, but this is not optimum. It might be because due to the all iterations are in the regime of $f(x_1, x_2) = \sqrt{x_1^2 + \gamma x_2^2}$ without changing the function. And for this function $(0, 0)$ is indeed global minimum. But in our case it is saddle point

3 Problem 3

Consider the problem of the form:

$$\min_{x} \sum_{i=1}^{N} f_i(1 - a_i^T x) + g(x)$$

where $g(x) = \frac{N\lambda}{2}\|x\|_2^2$

For consensus ADMM, we again reparametrize:

$$\min_{x_1,\ldots,x_N,x} \sum_{i=1}^{N} f_i(1 - a_i^T x_i) + g(x),$$
$$\text{subject to } x_i = x, i = 1, \ldots N$$

and this yields the decomposable ADMM updates:

$$x_i^{(k)} = arg\min_{x_i} f_i(1 - a_i^T x_i) + \frac{\rho}{2}\|x - x_i^{(k-1)} - w_i^{(k-1)}\|_2^2,$$

$$i = 1, \ldots, N$$

$$x_i^{(k)} = arg\min_{x_i} \frac{N\rho}{2}\|x - x_i^{(k)} - w_i^{(k-1)}\|_2^2 + g(x),$$

where $g(x) = \frac{N\lambda}{2}\|x\|_2^2$

$$w_i^{(k)} = w_i^{(k-1)} + x_i^{(k)} - x_i^{(k)},$$

$$i = 1, \ldots, N$$