Exercises for the course

August 12, 2018. A >= 7 points; B >= 4 points; C >= 2points

Problem 1 (2 points). One have to solve convex optimization problem ($\dim X = n^2$, $\sum_{i=1}^n \tilde{L}_i = \sum_{j=1}^n \tilde{W}_j = 1$)

$$f(X) = \sum_{i,j=1}^n c_{ij} X_{ij} + \gamma \sum_{i,j=1}^n X_{ij} \ln X_{ij} \to \min_{\substack{\sum_{j=1}^n X_{ij} = \tilde{L}_i,\ \sum_{i=1}^n X_{ij} = \tilde{W}_j \\ X_{ij} \geq 0,\ i,j=1,\dots,n;}} . \tag{1}$$

Show that (up to a sign) the dual problem will be ($\dim x = 2n$)

$$\varphi(x = (\lambda, \mu)) = \gamma \ln\left(\sum_{i,j=1}^n \exp\left(\frac{\lambda_i + \mu_j - c_{ij}}{\gamma}\right)\right) - \langle \lambda, \tilde{L} \rangle - \langle \mu, \tilde{W} \rangle \to \min_{x=(\lambda,\mu)\in\mathbb{R}^{2n}} . \tag{2}$$

Problem 2 (2 points). How to find such x^N that for the problem (2) $\varphi(x^N) - \varphi_* \leq \varepsilon$ with $N = \mathrm{O}\left(\sqrt{LR^2/\varepsilon}\right)$, $L = 2n/\gamma$, $R = \|x_*\|_2$ – 2-norm of the solution of the dual problem (2) and a cost of one iteration $\mathrm{O}(n^2)$?

Problem 3 (3 points). Propose a way to find such X^N that (see (1) and (2))

$$f(X^N) + \varphi(x^N) \leq \varepsilon, \ \sqrt{\sum_{i=1}^n \left(\sum_{j=1}^n X_{ij}^N - \tilde{L}_i\right)^2 + \sum_{j=1}^n \left(\sum_{i=1}^n X_{ij}^N - \tilde{W}_j\right)^2} \leq \tilde{\varepsilon},$$

with $N = \mathrm{O}\left(\sqrt{LR^2/\varepsilon}\right)$, $L = 2n/\gamma$, $R = \|x_*\|_2$ – 2-norm of the solution of the dual problem (2) and a cost of one iteration $\mathrm{O}(n^2)$.

Problem 4. (2 points) Let's consider QP-problem ($n \times n$ matrix $A \succ 0$ is fully completed, $|A_{ij}| \leq M$)

$$f(x) = \frac{1}{2}\langle x, Ax \rangle \to \min_{x \in S_n(1)} .$$

Using Fast Gradient Method of Yu. Nesterov, one can find ε-solutions for

$\mathrm{O}\left(n^2 \sqrt{M \ln n/\varepsilon}\right)$ arithmetic operations. // not good since $n \gg 1$ is huge

Show that if one use randomized Mirror Descend method with stochastic gradient $A^{\langle i[x] \rangle}$ – $i[x]$-column of matrix A and $P(i[x] = j) = x_j$, $j = 1,\dots,n$ (one can generate $i[x]$ for $\mathrm{O}(n)$ arithmetic operations), then one can find (ε, σ)-solution for $\mathrm{O}\left(nM^2 \ln n \cdot \ln(\sigma^{-1})/\varepsilon^2\right)$ arithmetic operations. Note, that vector $\bar{x}^N$ is a (ε,σ)-solution iff $f(\bar{x}^N) - f_* \leq \varepsilon$ with probability $\geq 1 - \sigma$.

Problem 5. (4 points) Assume that the optimal configuration determines by the solution of the convex problem

$$f(x) \to \min_{x \in Q}.$$

But each day (at each iteration) one can only observe independent stochastic gradients

$$\nabla_x f(x,\xi):\ E_\xi\left[\nabla_x f(x,\xi)\right] = \nabla f(x),\ \left\|\nabla_x f(x,\xi)\right\|_* \le M.$$

Mage can live $N \sim M^2R^2 \ln\left(\sigma^{-1}\right)/\varepsilon^2$ iterations and Expert $N \sim M^2R^2/\varepsilon^2$. Compare what is better to ask a solution from Mage or from $K \sim \ln\left(\sigma^{-1}\right)$ Experts $\bar{x}^K = \frac{1}{K}\sum_{i=1}^{K} \bar{x}^{N,i}$?

Problem 6 (2 points). Propose algorithm with complexity $\mathrm{O}\left(\varepsilon^{-1}\right)$ (required number of matrix-vector multiplications) that find ε-solution of the system $Ax = b$: $\left\|Ax^N - b\right\|_2 \le \varepsilon$. Try to estimate complexity more precisely.

Problem 7 (4 points). Assume that in Problem 6 $x_* \in S_n(1)$ (unit simplex in $\mathbb{R}^n$), $b = 0$ and square $n \times n$ matrix A has no more than $s \le \sqrt{n}$ nonzero elements at each row and each column ($n \gg 1$). Try to find ε-solution with complexity (total number of arithmetic operations) $\mathrm{O}\left(n + s^2 \ln(n)/\varepsilon^2\right)$.

Hint. Use conditional gradient method and try to explain how one can fulfill iteration for $\mathrm{O}\left(s^2 \ln(n)\right)$. For that one should keep gradient components in a heap with fast control to its maximal element.

Problem 8 (2 + 3 points). How one should solve

$$\frac{1}{2}\|Ax - b\|_2^2 + \mu \sum_{k=1}^{n} x_k \ln x_k \to \min_{\sum_{k=1}^{n} x_k = 1,\ x \ge 0} \ ?$$

In case a) $\mu \ge 0$ – is small ; b) μ – isn't small.

Problem 9 (3 + 2 points). Propose an efficient method for

$$\sum_{k=1}^{m} f_k\left(A_k^T x\right) + g(x) \to \min_{x \in Q},$$

where $g(x) = \frac{1}{2}\|x\|_2^2$ and

a) $f_k(y_k) = C \max\left\{0, 1 - b_k y_k\right\}$,

b) $f_k(y_k) = C \cdot \left(y_k - b_k\right)^2$.