Assignment 7

1. (**60**) In this problem we consider minimization of

$$f(x) = x^T A x + 2b^T x \tag{1}$$

under the constraint $x^T x \leq 1$. We introduce Lagrangian

$$L(x, \lambda) = x^T A x + 2b^T x + \lambda(\|x\|_2 - 1) \tag{2}$$

and consider both the primal problem

$$\min_{x,\ \|x\|_2 \leq 1} f(x) = \min_x \max_{\lambda \geq 0} L(x, \lambda) \tag{3}$$

and the dual problem (obtained by interchanging max and min operations):

$$\max_{\lambda \geq 0} \min_x L(x, \lambda) = \max_{\lambda \geq 0} g(\lambda), \quad \text{where} \quad g(\lambda) = \min_x L(x, \lambda) \tag{4}$$

- (**10**) For $A = \begin{pmatrix} -1 & 3 \\ 3 & -1 \end{pmatrix}$ and $b = (1, 1)$, solve the dual problem: compute explicitely $g(\lambda)$ and then $\max_{\lambda \geq 0} g(\lambda)$.
- (**10**) For the same A, b, solve the primal problem. In order to achieve this, write down the four KKT conditions for x and λ (stationarity, primal feasibility, dual feasibility and complementary slackness). How many solutions (λ, x) does this system of equations and inequalities have? Which of them corresponds to the global minimum?
- (**5**) What is the duality gap for this problem? In this exercise, you may quantify this gap for a particular instance of A, b but in fact it remains the same for arbitrary A, b.
- As we have seen above, some of the solutions to KKT conditions do not yield the global minimum of our problem. It turns out that adding one more equation is sufficient to pick the global minimum:

$$A + \lambda \hat{1} \succcurlyeq 0. \tag{5}$$

  Check that one of the solutions from KKT equations in the example above does not satisfy this equation (**5**). Show that, in general, this condition is indeed a necessary one (**10**).
- (**20**) Finally, our aim is to solve the system of KKT conditions + Eq. (5). Using the eigendecomposition of the matrix A, construct an efficient algorithm to compute the global minimum (it should reduce to line search over λ). How many solutions does this system have? You may want to test your implementation using the example considered above.

2. (**30**) In this problem, we will apply convex optimization (via `cvxpy`) to study time series. Consider the following autoregressive model:

$$x_t = \alpha_t x_{t-1} + \omega_t \tag{6}$$

where $\omega_t \sim \mathcal{N}(0, 1)$ are IID normal and α_t is time-dependent in the range $(-1, 1)$.

- (**5**) For $\alpha_t \equiv \alpha$, what is the variance $\sigma^2 = \langle x_t^2 \rangle$ and the autocorrelation $\frac{\langle x_t x_{t+\tau} \rangle}{\sigma^2}$ of the time series?
- (**10**) For the length of $T = 1000$, generate a piece-wise constant α_t with 10 switching points at random monents of time, so that $\alpha_t = (\alpha_0, \alpha_0, ...\alpha_0, \alpha_1, \alpha_1, ..., \alpha_9)$. With this choice of α_t, generate one sample of time series, according to Eq. 6. Optimizing the cost function

$$\sum_{t=1}^{T} \left[(x_t - \alpha_t x_{t-1})^2 + \gamma |\alpha_t - \alpha_{t-1}| \right] \to \min$$

  over α_t, try to recover α_t (your estimate will be certain $\hat{\alpha}_t$). What is the value of γ which gives an adequate result? (its straighforward to determine via experimentation since you know the ground truth). Check your result over several samples of α_t and corresponding x_t.
- (**15**) In the file `x.npy`, you fill find a longer time series generated in a similar way but with a different number of switch points. Estimate the number of switching points and α_t for this time series.