Technical University of Crete

School of Electrical and Computer Engineering

Course: Advanced Topics in Convex Optimization

Instructor: Athanasios P. Liavas

Exercises, Spring 2022.

New exercises will be added regularly.

You must try to implement as many as you can.

Most recent changes in red font.

1. Compute the projection onto the unit simplex (see Section 6.4.3 and Corollary 6.29). Check with CVX.

2. Compute a point of the set $S = S_1 \cap S_2$ (assuming that $S \neq \emptyset$), where

$$S_1 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}, \quad S_2 = \mathbb{R}^n_+. \tag{1}$$

   See Example 8.23. First, check with CVX that $S \neq \emptyset$ and then implement the algorithm.

3. Let $\mathbf{c} \in \mathbb{R}^n$ and $\Delta_n = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^n x_i = 1.\}$. Solve the KKT to solve the problem

$$\min_{\mathbf{x} \in \Delta_n} \mathbf{c}^T \mathbf{x}. \tag{2}$$

4. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ ($m \gg n$, $m \gtrapprox n$) and consider the problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1. \tag{3}$$

   Generate $\mathbf{A}$ and $\mathbf{b}$ using (1) Gaussian distribution and (2) uniform distribution.

   (a) Solve the problem via CVX and compute $\mathbf{x}^*$ and f_{opt}.

(b) Implement the subgradient descent algorithm with the Polyak step (see Example 3.44). As subgradient, use the "weak result" stated at the end of the example, i.e.

$$\mathbf{A}^T \mathrm{sgn}(\mathbf{A}\mathbf{x} - \mathbf{b}) \in \partial f(\mathbf{x}). \tag{4}$$

Terminating condition: `STOP` if

$$f(\mathbf{x}^k) \leq c \cdot f_{\mathrm{opt}}, \tag{5}$$

where $c \gtrapprox 1$ (for example, $c = 1.1, 1.01, 1.001$). Use `semilogy` and plot $f_{\mathrm{best}}^k - f_{\mathrm{opt}}$ versus k.

(c) Solve the problem with the subgradient algorithm with dynamic step-size (use the same terminating condition and produce the corresponding plot $f_{\mathrm{best}}^k - f_{\mathrm{opt}}$ versus k).

(d) Put the two plots in the same figure (experiment with $m \gg n$ and $m \gtrapprox n$ and Gaussian and uniform distributions). What do you observe?

(e) Compute L_f such that $\|\mathbf{g}\|_2 \leq L_f$ for all $\mathbf{g} \in \partial f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$ (see Assumption 8.12). Compute and plot the upper bound (see Theorem 8.13 (c))

$$f_{\mathrm{best}}^k - f_{\mathrm{opt}} \leq \frac{L_f \|\mathbf{x}^0 - \mathbf{x}^*\|_2}{\sqrt{k+1}}. \tag{6}$$

Plot the upper bound and the quantities ploted in question (d) in the same `semilogy`.

(f) Solve the problem with the stochastic gradient algorithm. Compute the cost function every epoch (1 epoch is equal to m stochastic gradient iterations), i.e. $f(\mathbf{x}^{km})$. Run the algorithm for as many epochs as the number of iterations required for convergence in step (c) and put in `semilogy` quantity $f(\mathbf{x}^{km}) - f_{\mathrm{opt}}$ versus k (k is the number of epochs).

(g) Solve the problem with the incremental algorithm. Run the algorithm for as many outer iterations as the number of iterations required for convergence in step (c) and produce the corresponding `semilogy`.

(h) Put all plots together.

(i) Suggested implementation of optimization algorithms:

```
[x_est, f_k] = alg_name(A, b, x_init, f_opt, c, step_size_type)
```

5. Repeat steps (a)-(e) of Exercise 4 for the cost function

$$f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_\infty. \tag{7}$$

6. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Generate *sparse* vector $\mathbf{x}_s \in \mathbb{R}^n$ (let $s \ll n$ be the number of nonzero elements of $\mathbf{x}_s$, and let $m \gtrapprox 2s\log(n) \ll n$). Compute $\mathbf{b} = \mathbf{A}\mathbf{x}_s$. Try to recover $\mathbf{x}_s$ as follows:

   (a) use CVX, solve the problems (set $\lambda = 0.01, 0.1, 1, 10$)

   $$\min_{\mathbf{x}} f_1(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1 \tag{8}$$

   and

   $$\min_{\mathbf{x}} f_2(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_2^2. \tag{9}$$

   and compute $\mathbf{x}^*$ and f_{opt}. Compare $\mathbf{x}^*$ and $\mathbf{x}_s$. What do you observe?

   (b) Solve problem (8) using the algorithms ISTA and FISTA. Terminating condition: $f_1(\mathbf{x}^k) < c \cdot f_{\text{opt}}$, for $c \gtrapprox 1$. Plot, in a common `semilogy`, quantity $f_1(\mathbf{x}^k) - f_{\text{opt}}$ versus k for the three algorithms. What do you observe?

7. Generate a piece-wise constant vector $\mathbf{x}_{\text{pwc}} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, and

$$\mathbf{b} = \mathbf{A}\mathbf{x}_{\text{pwc}} + \mathbf{e}. \tag{10}$$

   Try to estimate $\mathbf{x}_{\text{pwc}}$ by solving, e.g. via CVX, the following problems (perform experiments with weak and strong noise):

   (a) the simple LS problem $\min_{\mathbf{x}} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$;

   (b) the regularized LS problem

   $$\min_{\mathbf{x}} F(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\mathbf{D}\mathbf{x}\|_1, \tag{11}$$

   where $\mathbf{D} \in \mathbb{R}^{n,n}$ is given by

   $$\mathbf{D} = \rho \begin{bmatrix} +1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & +1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{12}$$

   with $\rho > 0$.

(c) Solve problem (11) using the S-FISTA algorithm, and plot, in a `semilogy`, quantity $f(\mathbf{x}^k) - f_{\text{opt}}$ versus k. Terminating condition: `STOP` if $f(\mathbf{x}^k) - f_{\text{opt}} \le \epsilon$, for small, positive ϵ (see Example 10.60).

(d) Solve problem (11) using the subgradient descent method, with dynamic step-size, and plot, in a `semilogy`, quantity $f(\mathbf{x}^k) - f_{\text{opt}}$ versus k. Run as many iterations as required in the previous step.

(e) Put both plots in the same figure. What do you observe?

8. Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} F(\mathbf{x}) = f(\mathbf{x}) + \lambda g(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1 + \lambda\|\mathbf{x}\|_1. \tag{13}$$

Solve the problem via CVX and compute $\mathbf{x}^*$ and f_{opt}. Solve the problem with (terminating condition for all algorithms: `STOP` if $f(\mathbf{x}^k) \le c \cdot f_{\text{opt}}$, for $c \gtrapprox 1$ – test the cases $m \gg n$, $m \approx n$, and $m \ll n$).

(a) the projected subgradient method (with both Polyak and dynamic step);

(b) the proximal subgradient method (see Example 9.29);

(c) Smooth function $f(\mathbf{x})$ and solve the resulting problem with the S-FISTA algorithm (see Example 10.60 for an "analogous" problem);

(d) Smooth both functions and solve the resulting problem with the accelerated gradient algorithm (S-FISTA with no proximal step).

9. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} h(\mathbf{x}) = \|\mathbf{x}\|_1, \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{14}$$

(a) Solve the problem using the CVX and compute $\mathbf{x}^*$ and h_{opt}.

(b) Compute a smooth approximation, h_μ, to h and solve the problem via the S-FISTA algorithm (see Example 10.59). Terminating condition: `STOP` if $h(\mathbf{x}^k) - f_{\text{opt}} < \epsilon$ (note that h is Lipschitz continuous, with constant $l_h = \sqrt{n}$).

(c) Solve the problem via the projected subgradient descent algorithm (run the algorithm for as many iterations as required in step (b) to converge).

(d) Plot in the same `semilogy` quantity $h(\mathbf{x}^k) - h_{\text{opt}}$, for the two algorithms. What do you observe?

10. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, and consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1, \text{ subject to } \mathbf{x} \in \Delta_n. \tag{15}$$

   (a) Solve the problem via CVX and compute $\mathbf{x}^*$ and f_{opt}.

   (b) Solve the problem via the projected subgradient descent. Terminating condition: `STOP` if $f(\mathbf{x}^k) \leq c \cdot f_{\text{opt}}$. Start the algorithm from $\mathbf{x}^0 = \frac{1}{n}\mathbf{1}$.

   (c) Solve the problem via the mirror descent using the same terminating condition (see Example 9.19). Start the algorithm from $\mathbf{x}^0 = \frac{1}{n}\mathbf{1}$.

   (d) Put in the same `semilogy` quantities $f(\mathbf{x}^k) - f_{\text{opt}}$ versus k, for both algorithms. What do you observe?

   (e) Experiment with (a) random $\mathbf{A}$ and $\mathbf{b}$ and (b) $\mathbf{b} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{e}$.

11. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ (with $m \ll n$), $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{c} \in \mathbb{R}^n$. Consider the linear program in standard form

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}, \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}. \tag{16}$$

   Generate a feasible problem, solve it via CVX, and compute $\mathbf{x}^*$ and f_{opt}.

   Express the problem in an equivalent form as follows:

$$\min_{\mathbf{x},\mathbf{z}\in\mathbb{R}^n} f(\mathbf{x}) + g(\mathbf{z}), \text{ subject to } \mathbf{x} = \mathbf{z}, \tag{17}$$

   where

$$g(\mathbf{z}) = \delta_{\mathbb{R}^n_+}(\mathbf{z}), \tag{18}$$

   and

$$\text{dom}(f) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}. \tag{19}$$

   Solve the problem via the ADMM algorithm, as follows:

   (a) write down the ADMM udpates,

   (b) solve the optimization subproblems,

   (c) implement the solutions efficiently,

   (d) plot in a `semilogy` diagram quantities $f(\mathbf{x}^k) - f_{\text{opt}}$.

12. Let $\mathbf{d}, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, with $\mathbf{v}_1 \leq \mathbf{v}_2$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, and $\mathbf{b} \in \mathbb{R}^m$. Compute the projection of $\mathbf{d}$ onto the following sets (make sure that they are nonempty):

(a) $\mathcal{S}_1 = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{x} \le \mathbf{b}\}$,

(b) $\mathcal{S}_2 = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{v}_1 \le \mathbf{x} \le \mathbf{v}_2\}$,

(c) $\mathcal{S}_3 = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{x} \le \mathbf{b}, \mathbf{v}_1 \le \mathbf{x} \le \mathbf{v}_2\}$.

13. Solve the total variation denoising problem (see Section 12.4.3) with various types of (noiseless) $\mathbf{x}$ (smooth, piece-wise constant), additive white Gaussian noise, and l_1 and l_2 regularizers.

    (a) solve the problems with CVX;

    (b) solve the l_1 problem using the DPG and the FDPG algorithms (the special structure of matrix $\mathbf{D}$ makes it possible to avoid *all* matrix-vector multiplications and use instead, carefully chosen, vector operations).