NUMERICAL OPTIMISATION
TUTORIAL 27/03/20
ASSIGNMENT 4 (submit by 11pm Thursday 16/04)

MARTA BETCKE

Submit solutions to exercises 2,4 and 5 to `TurnitIn` in a form of a single PDF report. No hand written solutions will be accepted. Include Matlab code for Augmented Lagrangian and Quadratic penalty methods in exercise 5. These codes should be no longer than about 1 page each and the entire submission should not exceed about 6 pages.

EXERCISE 1 [DEMO]

Consider a problem to minimise the function

$$\min_x f(x) = \frac{1}{2}x^T Gx + c^T x$$

subject to the constraint

$$Ax \leq b,$$

where $G \in \mathbb{R}^{n\times n}$ symmetric positive semidefinite, $A \in \mathbb{R}^{m\times n}, m \leq n$, full row rank, $c \in \mathbb{R}^n, b \in \mathbb{R}^m$.

(a) State the KKT conditions for this problem.

(b) Rewrite the constraint using a vector of slack variables $y \in \mathbb{R}^m, y \geq 0$ and give the corresponding KKT conditions.

(c) Formulate the dual to the problem in (b) and discuss its properties.

EXERCISE 2

Solve the following constraint minimisation problem:

$$\min_{(x,y)} f(x,y) = (x-2y)^2 + (x-2)^2, \quad x - y = 4.$$

(a) Formulate the KKT system. **[10pt]**

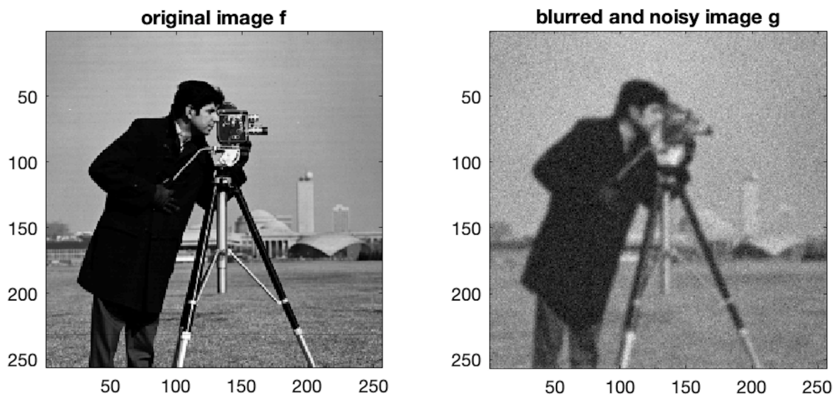
(b) Solve the KKT system (in any way you wish). Explain briefly your approach. **[10pt]**

EXERCISE 3 [DEMO]

The 256×256 Cameraman image $f \in \mathbb{R}^{n\times n}$ (left) was we blurred and contaminated with noise according to the linear model (right)

$$g = Af + \epsilon.$$

Here A is a discretised two-dimensional convolution operator with a Gaussian kernel with zero mean $\mu = 0$ and a standard deviation $\sigma = 1.5$ and ϵ denotes additive Gaussian white (identically distributed and uncorrelated) noise. .



The deblurring problem is to recover f from the blurred and noisy image g. The Total Variation (TV) has proven to be a valuable prior for recovery of piecewise constant or piecewise smooth images. This amounts to solving the following minimisation problem

$$f_{\text{rec}} = \arg\min_f \frac{1}{2}\|Af - g\|_2^2 + \alpha TV(f).$$

Here we are going to use the anisotropic total variation

$$TV(f) = \sum_{j=1}^{n}\sum_{j=1}^{n} |f_{(i+1,j)} - f_{(i,j)}| + |f_{(i,j+1)} - f_{(i,j)}|$$

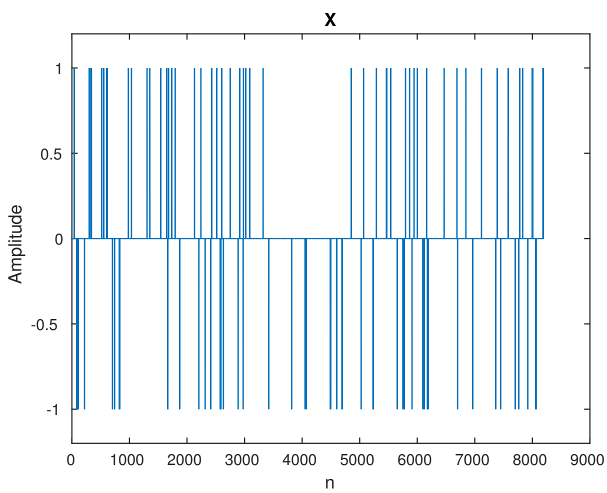
with 0 boundary conditions.
Solve the above minimisation problem with ISTA, FISTA and ADMM. Discuss the choices of parameters. Explain the convergence behaviour relating to the theory.

EXERCISE 4

Consider a *sparse* signal $x = \{x_n\}_{n=1\dots N}$ of length $N = 2^{13}$ consisting of

- $T = 100$ randomly distributed spikes with values $\{\pm 1\}$.
- the remaining, $N - T$, values equal to 0.

A possible realisation of the signal x is:



Following the compressed sensing paradigm we take $K = 2^{10}$ *compressed measurements* of the form

$$y_i = a_i^{\mathrm{T}} x,$$

where $a_i \in \mathbb{R}^N$ is an appropriately chosen sensing vector. Assuming that the measurement is corrupted with zero-mean Gaussian noise with standard deviation $\sigma = 0.005$, the measurements can be compactly written as

$$\tilde{y} = Ax + e,$$

where a_i^{T} is the ith row of the measurement matrix $A \in \mathbb{R}^{K \times N}$ and $e \in \mathcal{N}(0, \sigma)$ is the normally distributed noise vector.

We consider two different measurement types

- A is a Gaussian random matrix with orthonormal rows (use `randn()` and `orth()` to construct it);
- A is a subsampled Welsh-Hadamard transform (the forward and inverse WH transform in MATLAB can be called via `fwht()`, `ifwht()` and subsampling corresponds to randomly choosing K rows).

While the problem to recover x from $\tilde{y}$ is underdetermined, under assumption of *sparsity* it is still possible to recover sparse signals from such incomplete measurements (under certain assumptions on A and sparsity of the signal vs. number of measurements). While forcing sparsity would lead to a combinatorial problem, the L_1 norm provides a relaxation yielding solution to the original problem under mild assumptions. The resulting compressed sensing recovery problem reads

$$x_{CS} = \arg\min_x \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_1, \tag{CS}$$

where λ is the regularisation parameter chosen depending on the properties of A and the standard deviation of the noise.

Set up and solve the compressed sensing recovery problem (CS) with ISTA, FISTA and ADMM. Discuss the choice of parameters. Explain the convergence behaviour relating to the theory. **[40pt]**

EXERCISE 5

Consider the constraint optimisation problem

$$\begin{aligned} f(x, y) &= (x-a)^2 + \frac{1}{2}(y-b)^2 - 1 \\ \text{s.t.} \quad & x^2 + y^2 = 2 \end{aligned}$$

with $a = 1, b = 1.5$.

Implement a simple version of the quadratic penalty and augmented Lagrangian methods. Note that:

- Both formulations result in an unconstraint optimisation problem which can be solved by any method for unconstraint problems.
- The framework for both methods is very similar with two main differences
  - In penalty method the penalty parameter μ is increased while in augmented Lagrangian it can be chosen fixed (the method will be exact for parameter values above a certain threshold);
  - In augmented Lagrangian the Lagrange multiplier approximation is updated.
- Ignore the ill-conditioning of the Hessian.
- The decreasing tolerance in quadratic penalty method is not very sensitive as long as $\tau_k \to 0$.

Some suggestions to guide your implementation

- Use line search method with a Newton direction and backtracking line search to solve the unconstraint problem at each step.
- Use $\|x_k - x_{k-1}\| < \varepsilon$ as a stopping criterium for Augmented Lagrangian and Quadratic penalty methods. Set the final tolerance fairly small $\varepsilon = 1e-10$.
- A good choice of parameters are $\mu_0 = 1$ for the initial penalty weight in the quadratic penalty method and $\mu = 10, \nu_0 = 1$ (fixed penalty and initial Lagrange multiplier) for the augmented Lagrangian method.

Solve this constraint optimisation problem using a feasible and infeasible starting point. Compare performance of both methods in terms of convergence rates and the path traced by the iterates. Relate your statements to the theory. **[40pt]**

EXERCISE 6 [DEMO]

Consider the function

$$\begin{aligned} f(x, y) &= (x-a)^2 + \frac{1}{2}(y-b)^2 - 1 \\ \text{s.t.} \quad & x^2 + y^2 \leq 2 \end{aligned}$$

with $a = 1, b = 1.5$ and $a = 0.5, b = 0.25$.

Solve this optimisation problems with interior point methods: primal dual and barrier methods. Discuss the choice of the initialisation point. Plot convergence of both methods in terms of relevant quantities and relate it to the theory. **[0pt]**