Numerical Optimization Assignment 4 Tutorial

Bolin Pan & Marta Betcke

Department of Computer Science,
Centre for Medical Image Computing,
Centre for Inverse Problems,
University College London

27th March 2020

Exercise 1 (a)

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, $A \in \mathbb{R}^{m\times n}, c \in \mathbb{R}^n, b \in \mathbb{R}^m$.

(a) State the KKT conditions for this problem.
We can rewrite $Ax \leq b$ as $Ax - b \leq 0$. The Lagrangian for this problem is

$$\mathcal{L}(x, \lambda, \nu) = \frac{1}{2}x^T Gx + c^T x + \lambda^T (Ax - b),$$

therefore, the KKT conditions can be written as

$$\begin{gathered} \nabla_x \mathcal{L}(x, \lambda, \nu) = Gx + c + A^T \lambda = 0, \\ Ax - b \leq 0, \\ \lambda \geq 0, \\ \lambda_i \cdot [Ax - b]_i = 0 \quad i = 1 \ldots m. \end{gathered}$$

Exercise 1 (b)

Consider a problem to minimise the function

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

subject to the constraint

$$Ax \leq b,$$

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

(b) Rewrite the constraint using a vector of slack variables $y \in \mathbb{R}^m, y \geq 0$ and give the corresponding KKT conditions.
We set $Ax - b + y = 0$ and $y \geq 0$. The Lagrangian for this problem is

$$\mathcal{L}(x, \lambda, \nu) = \frac{1}{2}x^T G x + c^T x - \lambda^T y + \nu^T (Ax - b + y)$$

and the KKT conditions can be expressed as

$$\begin{gathered}
\nabla_x \mathcal{L}(x, \lambda, \nu) = Gx + c + A^T \nu = 0, \\
\nabla_y \mathcal{L}(x, \lambda, \nu) = \lambda - \nu = 0 \Rightarrow \lambda = \nu, \\
Ax - b + y = 0, \\
y \geq 0, \\
\lambda \geq 0, \\
\lambda_i \cdot y_i = 0 \quad i = 1 \ldots m.
\end{gathered}$$

Exercise 1 (c)

$$\mathcal{L}(x,\lambda,\nu) = \frac{1}{2}x^TGx + c^Tx - \lambda^Ty + \nu^T(Ax - b + y)$$

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

Since the gradient of the Lagrangian for (b) is

$$\nabla_x \mathcal{L}(x,\lambda,\nu) = Gx + c + A^T\nu,$$

and $\lambda = \nu$ (according to the KKT conditions) we find the dual for problem (a) with Lagrangian

$$\nabla_x \mathcal{L}(x,\lambda,\nu) = Gx + c + A^T\lambda.$$

This function has a unique zero that corresponds to the minimum of the quadratic form

$$x^* = -G^{-1}(A^T\lambda + c).$$

We substitute it into the Lagrangian to obtain the dual problem:

$$\begin{aligned}
\mathcal{L}(x^*,\lambda,\nu) &= \frac{1}{2}[-G^{-1}(A^T\lambda + c)]^T G[-G^{-1}(A^T\lambda + c)] + c^T(-G^{-1}(A^T\lambda + c)) + \lambda^T[A(-G^{-1}(A^T\lambda + c)) - b] = \\
&= \frac{1}{2}(A^T\lambda + c)^T G^{-1}(A^T\lambda + c) - (A^T\lambda + c)^T G^{-1}(A^T\lambda + c) - \lambda^T b = \\
&= -\frac{1}{2}(A^T\lambda + c)^T G^{-1}(A^T\lambda + c) - \lambda^T b
\end{aligned}$$

Therefore the dual problem is

$$\max_{\lambda} -\frac{1}{2}(A^T\lambda + c)^T G^{-1}(A^T\lambda + c) - \lambda^T b$$

$$\text{subject to } \lambda \geq 0$$

The Lagrangian for (b) is the same as for (a).

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]**

Hints:

a) Try Exercise 1

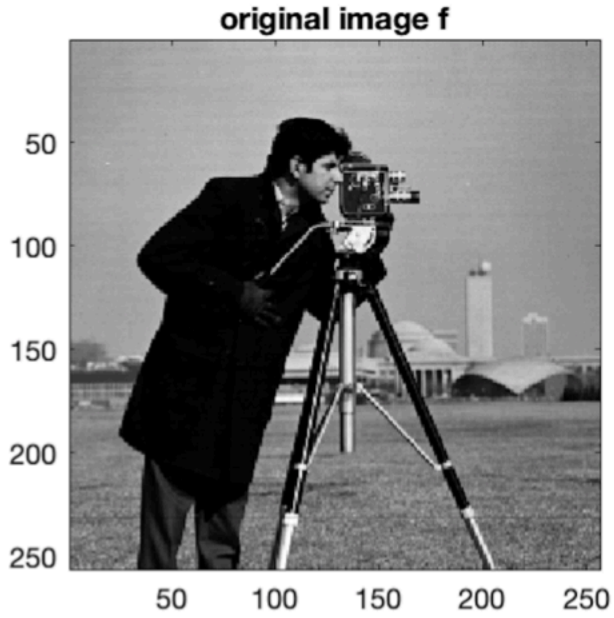
b) The KKT system can be simply solved by hand

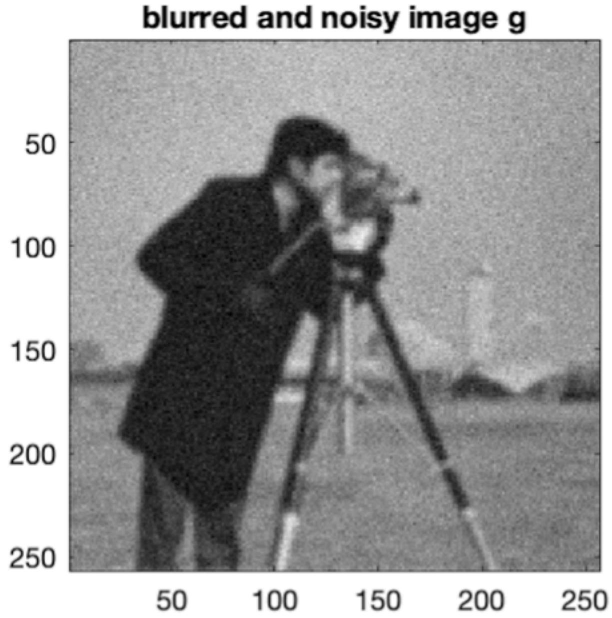
Exercise 3

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. .





Exercise 3

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. .

- **Task - Inverse Problems**

  Assume f is unknown, try to recover f from noisy image g
  - Denoising (additive noise)
  - Deblurring (gaussian filter)

- **Simple Approaches**

  - Direct inversion: $f = A^{-1}g$ (ill-posed)
  - Least squares: $\min \frac{1}{2}||g - Af||_2^2$ (no prior knowledge, equivalent to minimise likelihood)

Exercise 3

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. .

- **Variational Regularization**

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

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

- Prior - The Total Variation (TV term)
- TV is a valuable prior for recovery of piecewise constant or piecewise smooth images.

Exercise 3

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

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

- **ISTA**

$$\begin{aligned} x_k &= \text{prox}_{\tau_k g}(x_{k-1} - \tau_k \nabla f(x_{k-1})) \quad \textbf{Prox of } \boldsymbol{g} \textbf{ - PDHG} \\ &= \arg\min_x \left\{ g(x) + \frac{1}{2\tau_k}\|x - (x_{k-1} - \tau_k \nabla f(x_{k-1}))\|^2 \right\}. \end{aligned}$$

$\tau_k \in (0, 2/L(f))$ **Lipschitz constant**

- **FISTA**

Initialize: $y_1 := x_0 \in \mathbb{E}, \ \tau_1 = 1.$

Step k :

$$x_k = \text{prox}_{1/L}(g)\left(y_k - \frac{1}{L}\nabla f(y_k)\right)$$

$$\tau_{k+1} = \frac{1 + \sqrt{1 + 4\tau_k^2}}{2}$$

$$y_{k+1} = x_k + \frac{\tau_k - 1}{\tau_{k+1}}(x_k - x_{k-1}).$$

Exercise 3

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

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

- **ADMM** $$\min_{x\in\mathbb{R}^n,\, z\in\mathbb{R}^m} f(x) + g(z) \quad \text{subject to } Ax + Bz = c \qquad \text{(SCOP)}$$

$$x_{k+1} = \arg\min_x f(x) + \rho/2\|Ax + Bz_k - c + u_k\|_2^2$$

Prox of *g* - PDHG $$z_{k+1} = \arg\min_z g(z) + \rho/2\|Ax_{k+1} + Bz - c + u_k\|_2^2$$

$$u_{k+1} = u_k + \underbrace{Ax_{k+1} + Bz_{k+1} - c}_{r_{k+1}}.$$

- **SCOP:** $$f(x) = \frac{1}{2}||Ax - g||_2^2$$

$$g(z) = TV(z)$$

Subject to $x - z = 0$

Exercise 3 ADMM Acceleration

- **ADMM**

$$x_{k+1} = \arg\min_x f(x) + \rho/2\|Ax + Bz_k - c + u_k\|_2^2$$
$$z_{k+1} = \arg\min_z g(z) + \rho/2\|Ax_{k+1} + Bz - c + u_k\|_2^2$$
$$u_{k+1} = u_k + \underbrace{Ax_{k+1} + Bz_{k+1} - c}_{r_{k+1}}.$$

1. Varying Penalty Parameter

$$\rho^{k+1} := \begin{cases} \tau^{\text{incr}}\rho^k & \text{if } \|r^k\|_2 > \mu\|s^k\|_2 \\ \rho^k/\tau^{\text{decr}} & \text{if } \|s^k\|_2 > \mu\|r^k\|_2 \\ \rho^k & \text{otherwise,} \end{cases}$$

$$s^{k+1} = \rho A^T B(z^{k+1} - z^k)$$
$$r^{k+1} = Ax^{k+1} + Bz^{k+1} - c$$

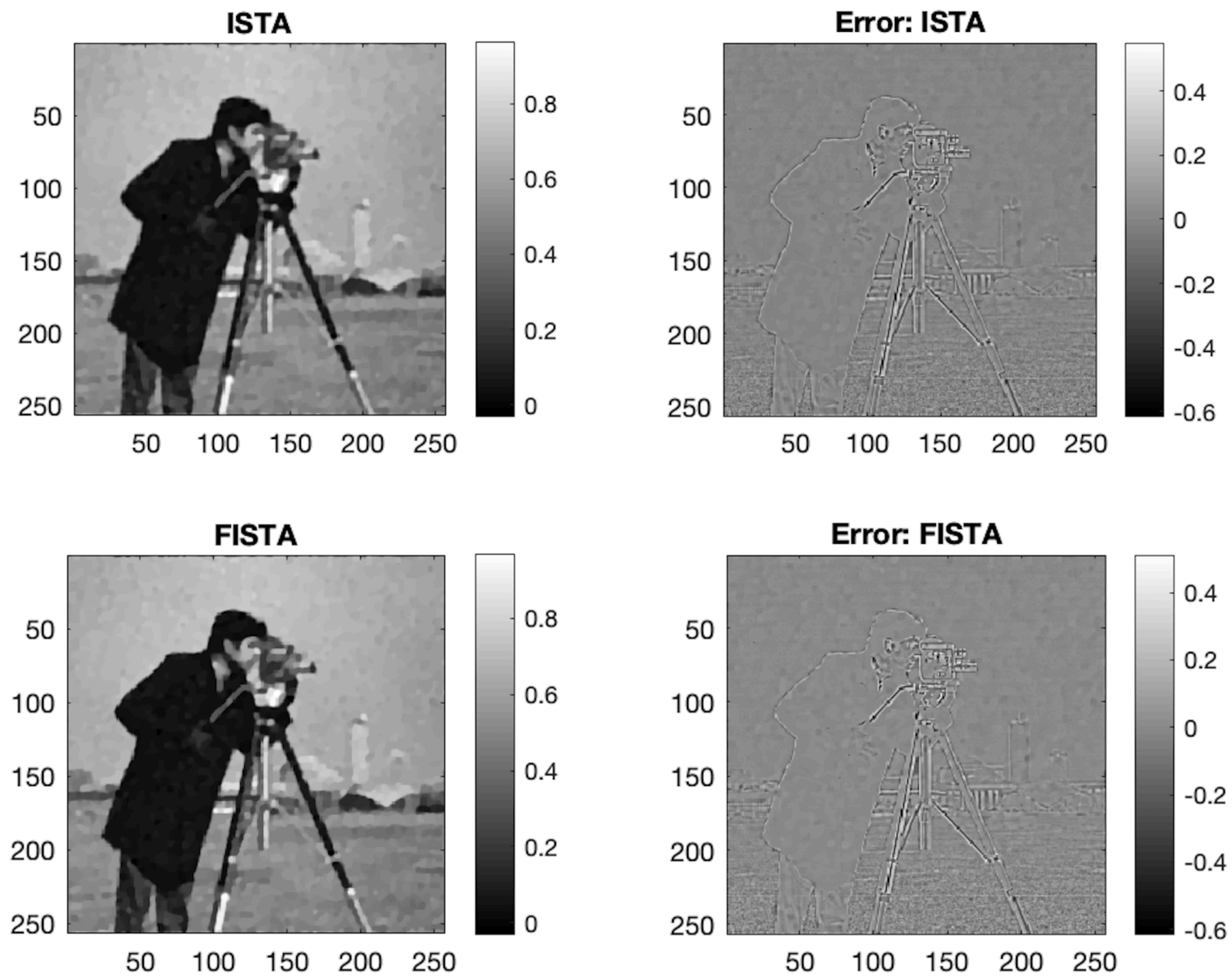
2. Over-relaxation – In z and u updates

$$\alpha^k Ax^{k+1} - (1-\alpha^k)(Bz^k - c), \qquad \alpha^k \in [1.5, 1.8]$$

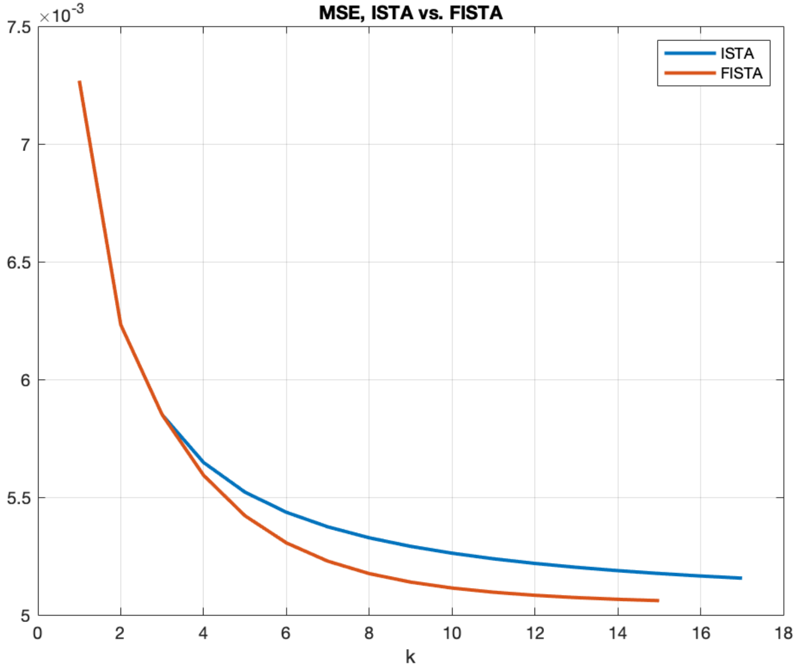
3. Stopping Criteria $\|r^k\|_2 \leq \epsilon^{\text{pri}}$ and $\|s^k\|_2 \leq \epsilon^{\text{dual}}$

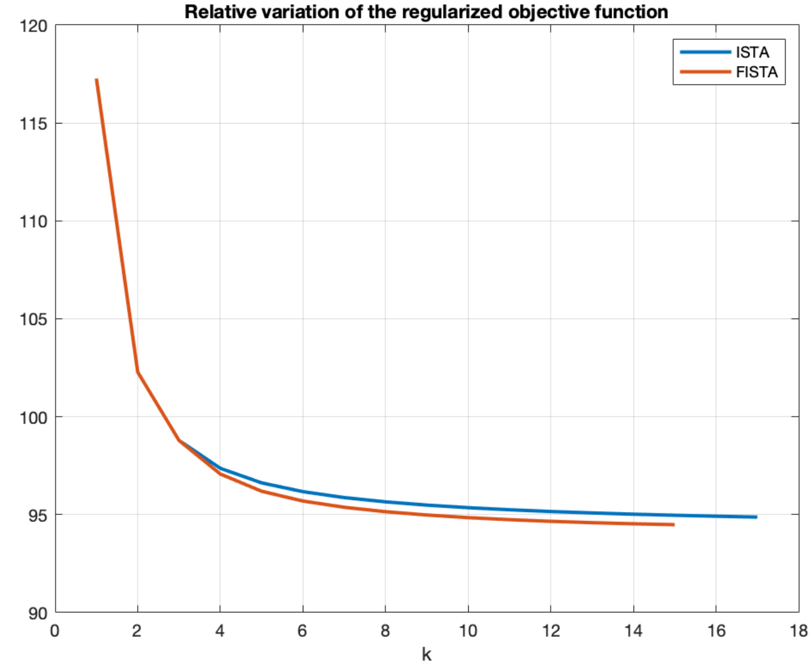
$$\epsilon^{\text{pri}} = \sqrt{p}\,\epsilon^{\text{abs}} + \epsilon^{\text{rel}}\max\{\|Ax^k\|_2, \|Bz^k\|_2, \|c\|_2\},$$
$$\epsilon^{\text{dual}} = \sqrt{n}\,\epsilon^{\text{abs}} + \epsilon^{\text{rel}}\|A^T y^k\|_2,$$

Exercise 3 ISTA vs. FISTA Reconstructions

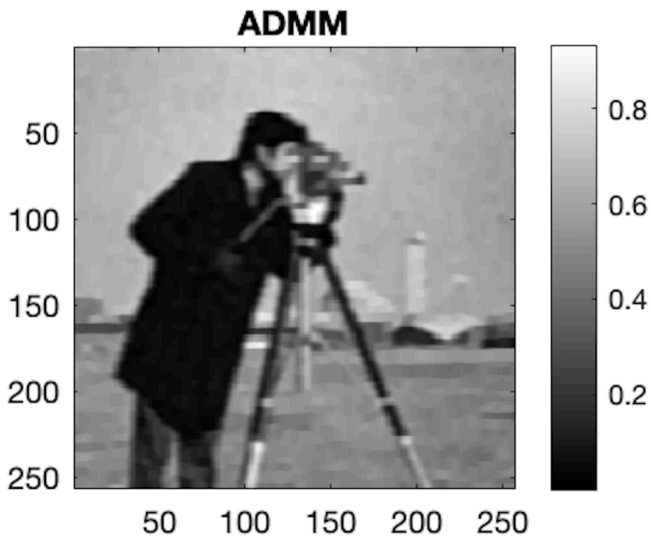


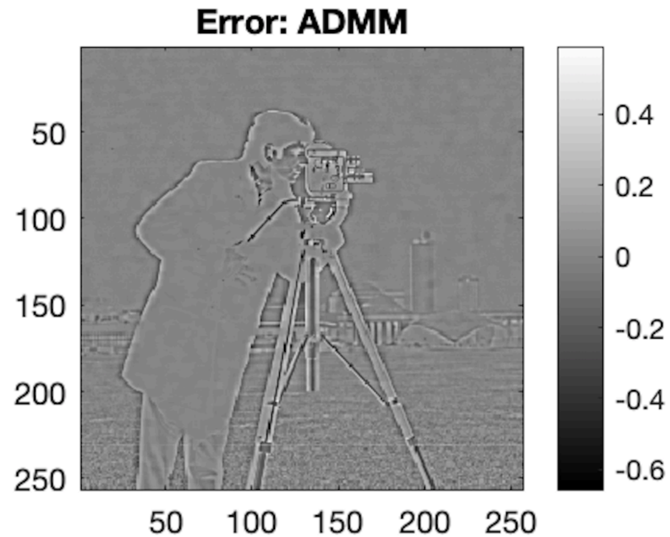
Exercise 3 ISTA vs. FISTA Convergence

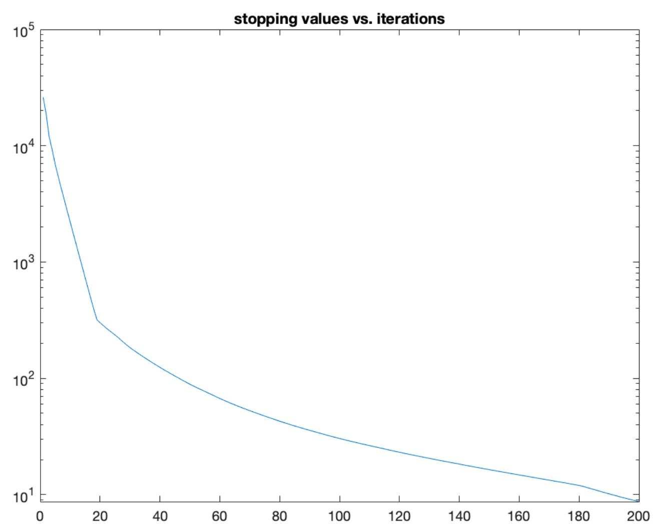




Exercise 3 ADMM Reconstruction







Exercise 4

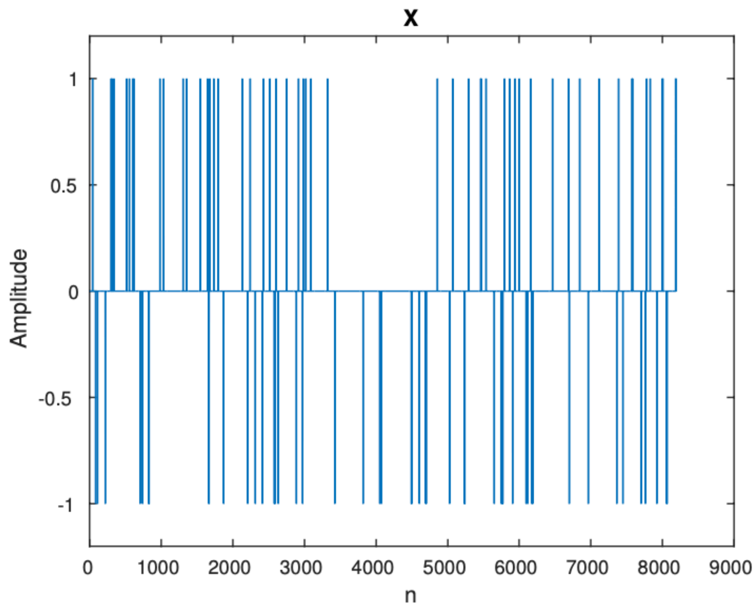
$$\tilde{y} = Ax + e,$$

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).

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



Exercise 4

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

- **FISTA & ISTA**

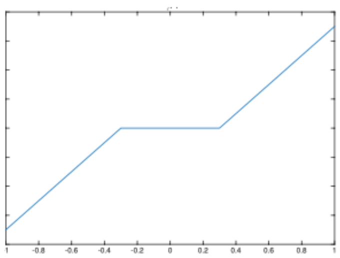
  1. Lipschitz constant = 1 as it is an CS problem
  2. Proximal operator

     - $f(x) = \|x\|_1$

       $$\mathrm{prox}_{\lambda f}(v) = S_\lambda(v),$$

       with elementwise soft thresholding

       $$S_\delta(x) = \begin{cases} x - \delta & x > \delta \\ 0 & x \in [-\delta, \delta] \\ x + \delta & x < -\delta \end{cases}$$



  3. Provide reconstructions, MSE, Convergent analysis

- **ADMM**

  1. Splitting
  2. Parameter tuning for update and accelerations
  3. Provide reconstructions, MSE, convergent analysis

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.

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

Consider the function

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

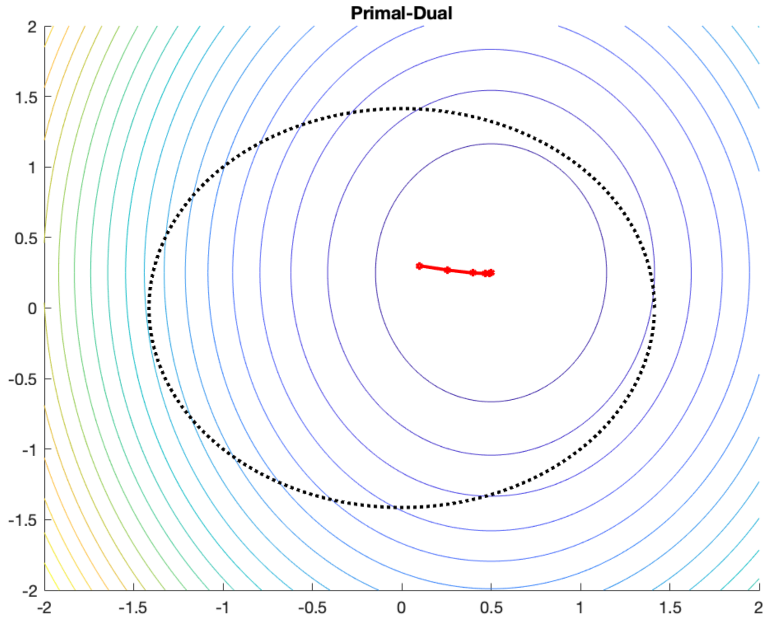
with $a = 1, b = 1.5$ and $a = 0.5, b = 0.25$.

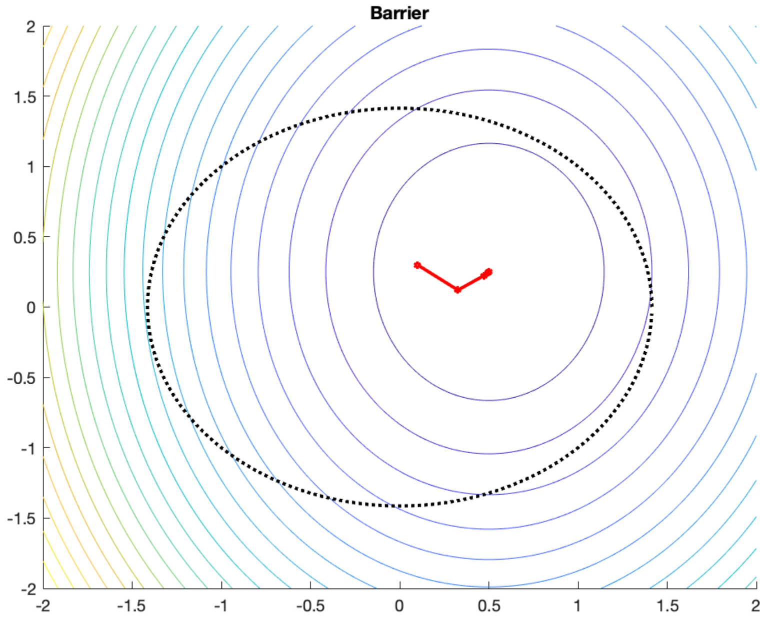
Solve this optimisation problems with interior point methods: primal dual and barrier methods. Discuss the choice of the initialisation point. Plot convergence of both methods in terms of relevant quantities and relate it to the theory. **[0pt]**

Exercise 6

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

- Strictly Feasible $a = 0.5, b = 0.25.$

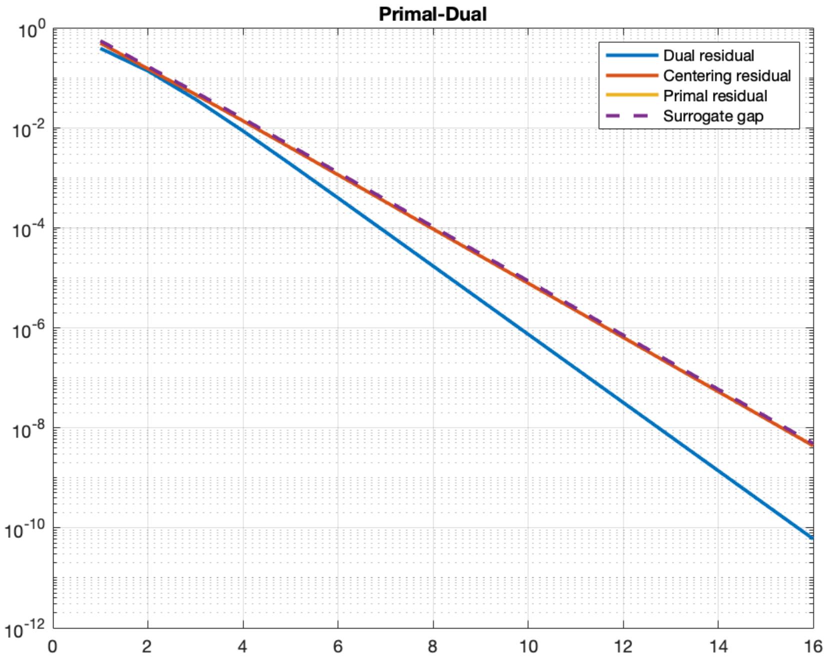


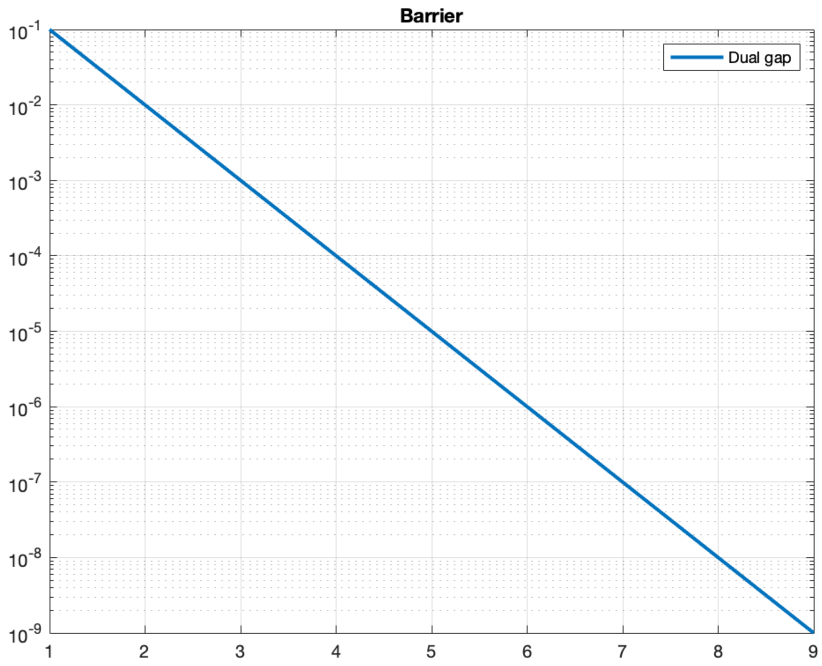


Exercise 6

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

- Strictly Feasible $a = 0.5, b = 0.25.$

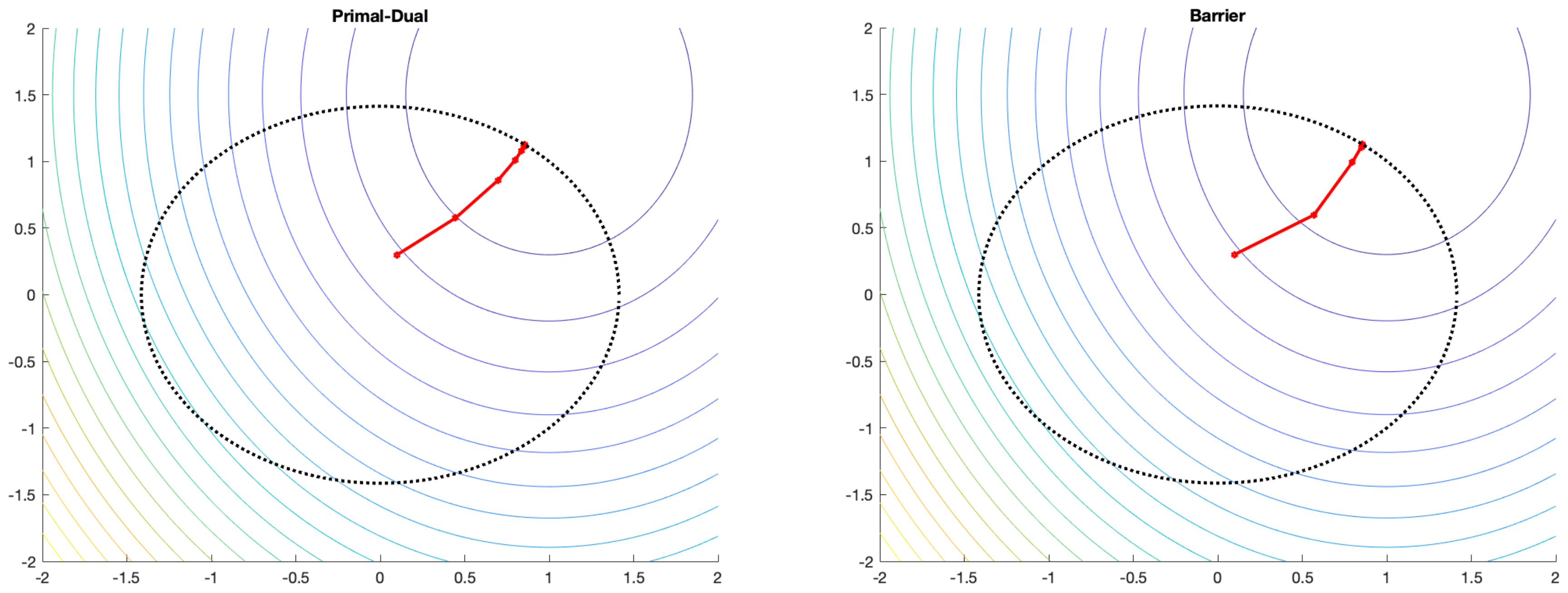




Exercise 6

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

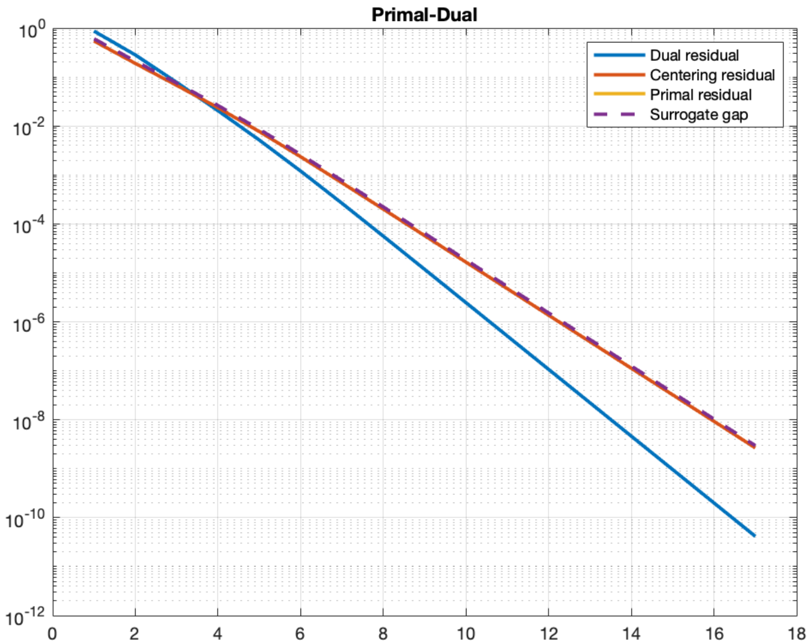
- **Feasible but not Strictly Feasible** $a = 1, b = 1.5$

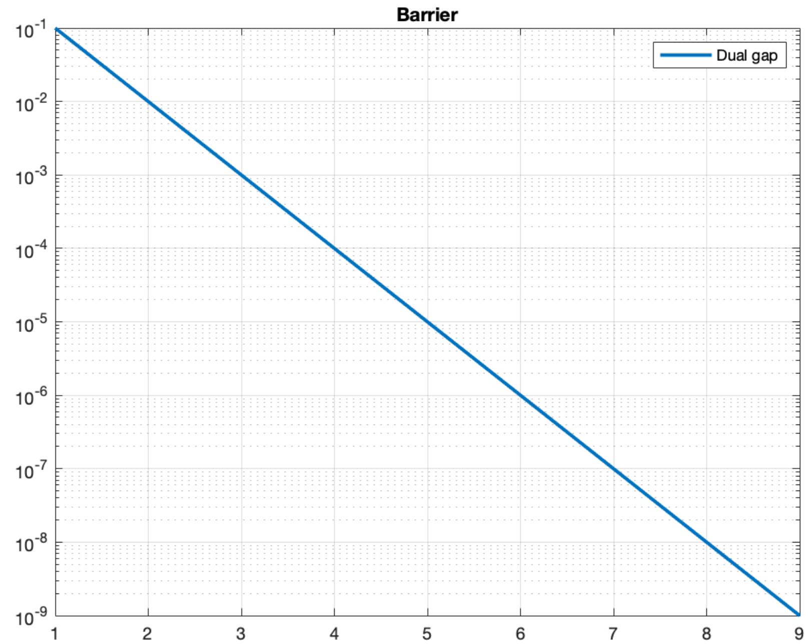


Exercise 6

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

- **Feasible but not Strictly Feasible** $a = 1, b = 1.5$





Thank You