Elements of Mathematics

Exercise Sheet 13

Submission due date: **08.02.21, 10:15h**

THEORY

All points of this sheet are bonus points!

1 Differentiable implies Continuous

Let $D \subset \mathbb{R}^n$, $x_0 \in D$ with $B_\varepsilon(x_0) \subset D$ for some $\varepsilon > 0$. Let $f: D \to \mathbb{R}$ be (Frechét-) differentiable at x_0. Show that f is continuous at x_0. *(4 Points)*

Solution:

Let $f: D \to \mathbb{R}^m$ be differentiable at x_0 with $A := Df(x_0)$ linear

$\Rightarrow \forall\, h_n \to 0,\ x_0 + h_n \in B_\varepsilon(x_0)$: $\frac{1}{\|h_n\|} \| f(x_0+h_n) - (f(x_0) + Ah_n) \|_2 \longrightarrow 0$

$\Rightarrow \forall$ —— " —— : $\| f(x_0+h_n) - (f(x_0) + Ah_n) \|_2 \longrightarrow 0$

$\overset{\Delta-\neq}{\Rightarrow} \forall$ —— " —— : $\| f(x_0+h_n) - f(x_0) \|_2 \leq \underbrace{\| f(x_0+h_n) - (f(x_0) + Ah_n) \|_2}_{\longrightarrow 0} + \underbrace{\| Ah_n \|_2}_{\longrightarrow 0}$

f differentiable

linear functions are continuous at 0

2 Derivatives

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix and $b \in \mathbb{R}^n$ a vector . Show that the function

$$f : \mathbb{R}^n \to \mathbb{R}, x \mapsto \tfrac{1}{2} x^T A x - b^T x$$

is Frechét differentiable and determine the gradient $\nabla f(x)$ of f at a point $x \in \mathbb{R}^n$.

Hint: Compute the directional (Gâteau) derivative and use the resulting expression as a candidate for the Fréchet derivative. *(4 Points)*

Solution:

- We first compute the Gâteaux derivative: Let $v \in \mathbb{R}^n$

$$\frac{1}{t}\left(f(x+tv) - f(x)\right) = \frac{1}{t}\Big[\tfrac{1}{2}(x+tv)^T A (x+tv) - b^T(x+tv) - x^T A x + b^T x\Big]$$

$$= \frac{1}{t}\Big[\cancel{\tfrac{1}{2}x^T A x} + \tfrac{2}{2} t\, x^T A v + \tfrac{1}{2} t^2 v^T A v - \cancel{b^T x} - t\, b^T v - \cancel{\tfrac{1}{2}x^T A x} + \cancel{b^T x}\Big]$$

$$= x^T A v - b^T v + \frac{t}{2} v^T A v$$

$$\xrightarrow{t \to 0} (Ax - b)^T v$$

$\Rightarrow$ Gâteau differentiable

- Now let us use $Df(x_0) := (Ax_0 - b)^T(\cdot)$ as candidate for the Fréchet derivative:

Let $x_0 \in \mathbb{R}^n$, $\varepsilon = 0$, $h_n \to 0$

$$\frac{1}{\|h_n\|} \| f(x_0 + h_n) - (f(x_0) + Df(x_0)(h_n)) \|_2$$

$$= \frac{1}{\|h_n\|} \Big| \tfrac{1}{2}(x_0 + h_n)^T A (x_0 + h_n) - b^T(x_0 + h_n) - \tfrac{1}{2}x_0^T A x_0 + b^T x_0 - (Ax_0 - b)^T h_n \Big|$$

$$= \frac{1}{\|h_n\|} \Big| \cancel{\tfrac{1}{2} x_0^T A x_0} + \cancel{x_0^T A h_n} + \tfrac{1}{2} h_n^T A h_n - \cancel{b^T x_0} - \cancel{b^T h_n} - \cancel{\tfrac{1}{2} x_0^T A x_0} + \cancel{b^T x_0} - \cancel{x_0^T A h_n} + \cancel{b^T h_n} \Big|$$

$$= \frac{1}{2} \frac{1}{\|h_n\|} \underbrace{|h_n^T A h_n|}_{\leq \|h_n\| \cdot \|Ah_n\|} \leq \frac{1}{2} \|Ah_n\| \xrightarrow{n \to \infty} 0$$

C-S-U

- To find the gradient we can evaluate $Df(x)(\cdot)$ on the standard basis:

$$\nabla f(x) = \begin{pmatrix} Df(x)(e_1) \\ \vdots \\ Df(x)(e_n) \end{pmatrix} = Ax - b$$

REMARK:

$A \in \mathbb{R}^{n \times n}$ symmetric and positive definite

$\Rightarrow$ $\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T A x - b^T x$ has an unique global minimum

Furthermore:

x^* is the unique minimizer of $f(x) := \frac{1}{2} x^T A x - b^T x$ $\Leftrightarrow$ $0 \overset{!}{=} \nabla f(x^*) = Ax^* - b$

f strictly convex

3 Heron's Method as Newton's Method

Consider the function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) := x^2 - a$ for some nonnegative number $a \geq 0$. Apply Newton's method to the nonlinear system $f(x) = 0$ (root finding problem). Compare the resulting iterative scheme to Heron's algorithm from earlier sheets. *(4 Points)*

Solution:

Idea of Newton's method: [see result in the script]

Let $f: \mathbb{R}^n \to \mathbb{R}^n$ be "sufficiently smooth". We want to find a root $\hat{x} \in \mathbb{R}^n$ of f such that $f(\hat{x}) = 0$.

We try to do this iteratively. Therefore let $x^k \in \mathbb{R}^n$ be our approximation to $\hat{x}$ at step k.

How to choose a step Δx^k such that the iteration

$$x^{k+1} := x^k + \Delta x^k$$

converges to $\hat{x}$?

By Taylor approximation:

$$f(x^{k+1}) \approx f(x^k) + J_f(x^k)\Delta x^k \overset{!}{=} 0 \qquad \text{for } \|\Delta x^k\| \text{ small}$$

suggest to take

$$\Delta x^k := -J_f^{-1}(x^k)\,\Delta x^k$$

such that all in all:

$$x^{k+1} := x^k + \Delta x^k = x^k - J_f^{-1}(x^k)\, f(x^k)$$

Example: $f: \mathbb{R}^1 \to \mathbb{R}^1$, $f(x) := x^2 - a$, $a > 0$

$$f'(x) = 2x, \quad (f'(x))^{-1} = \frac{1}{2x}, \quad x \neq 0$$

$$\Rightarrow x_{k+1} = x_k - (f'(x_k))^{-1} f(x_k)$$

$$= x_k - \frac{1}{2x_k}(x_k^2 - a)$$

$$= \frac{1}{2}\left(x_k + \frac{a}{x_k}\right)$$

4 Vector Space of Polynomials

Let $\mathbb{F}$ be a field. A set V together with a mapping $+$ (sum) and a mapping $\cdot$ (scalar multiplication) with

$$\begin{aligned} +: V \times V &\to V \qquad & \cdot: \mathbb{F} \times V &\to V \\ (v,w) &\mapsto v+w & (\lambda, v) &\mapsto \lambda \cdot v \end{aligned}$$

is called ***vector space (or linear space)*** *over* $\mathbb{F}$, if the following axioms **VR1** and **VR2** hold:

VR1 $(V,+)$ is a commutative (or abelian) group with neutral element 0, i.e.,

- **G1** Associativity: $\forall v_1, v_2, v_3 \in V: \; v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$
- **G2** Neutral element: $\forall v \in V: \; v + 0 = v$
- **G3** Inverse element: $\forall v \in V \; \exists_1 (-v) \in V: \; v + (-v) = 0$
- **G4** Commutativity: $\forall v_1, v_2 \in V: \; v_1 + v_2 = v_2 + v_1$

VR2 The scalar multiplication is consistent/compatible with $(V,+)$ in the following way:

for $\lambda, \mu \in \mathbb{F}$, $v, w \in V$ it holds that

(i) $(\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v$

(ii) $\lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w$

(iii) $\lambda \cdot (\mu \cdot v) = (\lambda \cdot \mu) \cdot v$

(iv) $1 \cdot v = v$

Furthermore, let $v_1, \dots, v_n \in V$, then with the summation and scalar multiplication we can more generally define the **span** as

$$\text{span}(v_1, \dots, v_n) = \{\sum_{i=1}^{n} \lambda_i v_i : \lambda_i \in \mathbb{F}\}.$$

Further we say that $v_1, \dots, v_n \in V$ are **linearly independent** if

$$\sum_{i=1}^{n} \lambda_i v_i = 0 \quad \Rightarrow \quad \lambda_i = 0 \;\; \forall\, i.$$

If $v_1, \dots, v_n \in V$ are linearly independent and $\text{span}(v_1, \dots, v_n) = V$, then we call $v_1, \dots, v_n$ a **basis of** V.
A mapping $f\colon V_1 \to V_2$ between two vector spaces is called **linear**, if

$$f(\lambda \cdot_1 v +_1 w) = \lambda \cdot_2 f(v) +_2 f(w)$$

for all $v, w \in V$. Here $+_1, \cdot_1$ and $+_2, \cdot_2$ denote the summation and scalar multiplication defined on V_1 and V_2, respectively. Examples are $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$ with the usual vector/matrix sum "$+$" and scalar multiplication "$\cdot$".
Now we consider another example of a vector space: Let $n \in \mathbb{N}$ and $P_n(\mathbb{R})$ be the set of all polynomials of degree $\leq n$ on $\mathbb{R}$, i.e., the set $P_n(\mathbb{R})$ contains all functions $p: \mathbb{R} \to \mathbb{R}$ of the form

$$p(x) = \sum_{k=0}^{n} \alpha_k x^k$$

for some $\alpha_0, \dots, \alpha_n \in \mathbb{R}$. We define a summation and scalar multiplication:

$$\begin{aligned} +&\colon P_n(\mathbb{R}) \times P_n(\mathbb{R}) \to P_n(\mathbb{R}), \; && (p+q)(x) := p(x) + q(x), \\ \cdot&\colon \mathbb{R} \times P_n(\mathbb{R}) \to P_n(\mathbb{R}), && (r \cdot p)(x) := r \cdot p(x). \end{aligned}$$

1. **VR axioms:** Please show that $P_n(\mathbb{R})$ together with the above defined summation and scalar multiplication forms a vector space.

   Hint: Check **VR1** and **VR2** with $V = P_n(\mathbb{R})$.

2. Let $k < m \in \mathbb{N}$. Compute

$$\lim_{x\to\infty} \frac{x^k}{x^m}, \quad \text{and} \quad \lim_{x\to\infty} \frac{\sum_{k=0}^{m-1} \alpha_k x^k}{x^m}$$

for arbitrary $\alpha_0, \ldots, \alpha_m \in \mathbb{R}$.

3. **Monomials form a basis:** Please show that the set $B := \{q_0, \ldots, q_n\}$ with

$$q_k : \mathbb{R} \to \mathbb{R}, \;\; x \mapsto x^k,$$

is a basis of $P_n(\mathbb{R})$. What is the dimension of the vector space $P_n(\mathbb{R})$?

Hint: Part (ii) basically provides the proof of linear independence and the other assertion is obvious from the definition of p.

4. **Derivative as linear operator:** Show that the operator $\mathcal{D} : P_n(\mathbb{R}) \to P_n(\mathbb{R}), \;\; p \mapsto p'$, which maps a polynomial to its first derivative, is a $\mathbb{R}$-linear function.

Hint: For $p(x) = \sum_{k=0}^n \alpha_k x^k$ we have $p'(x) = \sum_{k=0}^n \alpha_k k x^{k-1}$.

5. **Matrix representation of the derivative:** Let Φ be the linear, invertible function which maps a polynomial to its coefficients (coordinates in the above basis), i.e.,

$$\Phi : P_n(\mathbb{R}) \to \mathbb{R}^{n+1}, \;\; \sum_{k=0}^n \alpha_k x^k \mapsto (\alpha_0, \ldots, \alpha_n).$$

Please remark shortly why Φ is bijective. What is the matrix representation of the linear function $F : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ defined by

$$F := \Phi \circ \mathcal{D} \circ \Phi^{-1}$$

with respect to the standard basis $\{e_1, \ldots, e_{n+1}\}$? More precisely, derive the matrix $A \in \mathbb{R}^{(n+1)\times(n+1)}$ defined by

$$A := \begin{pmatrix} | & & | \\ F(e_1) & \ldots & F(e_{n+1}) \\ | & & | \end{pmatrix}.$$

(10 Points)

Solution:

1. **VR1:** Show, that $(P_n(\mathbb{R}), +)$ is an abelian group.
Let $p(x) = \sum_{k=0}^n \alpha_k x^k$, $q(x) = \sum_{k=0}^n \beta_k x^k$ and $w(x) = \sum_{k=0}^n \gamma_k x^k$ be in $P_n(\mathbb{R})$.

   (i) Associativity:

$$((p+q)+w)(x) = \sum[(\alpha_k + \beta_k) + \gamma_k]x^k = \sum[\alpha_k + (\beta_k + \gamma_k)]x^k = (p+(q+w))(x)$$

   (ii) Neutral Element:
   $0 := \sum_{k=0}^n 0 \cdot x^k$, then $\forall \; p \in P_n(\mathbb{R})$:

$$(0+p)(x) = \sum(0+\alpha_k)x^k = p(x)$$

   (iii) Inverse element:
   For $p(x) = \sum \alpha_k x^k$ define $-p(x) := \sum(-\alpha_k)x^k$,

$$\Rightarrow \quad (p+(-p))(x) = 0.$$

   (iv) Commutativity:

$$(p+q)(x) = \sum \underbrace{(\alpha_k + \beta_k)}_{=\beta_k+\alpha_k} x^k = (q+p)(x)$$

VR2: Consistency properties: Let $r, s \in \mathbb{R}$.

(i) $((r+s)p)(x) = \sum \underbrace{(r+s)\alpha_k}_{=r\alpha_k+s\alpha_k} x^k = (rp)(x) + (sp)(x)$

(ii)-(iv) ✓

2. Let $k < m \in \mathbb{N}$. Then

$$\begin{aligned} \frac{x^k}{x^m} &= x^{k-m} = \frac{1}{x^{m-k}} \overset{x\to+\infty}{\longrightarrow} 0 \\ \Rightarrow \quad \forall \alpha_0,\dots,\alpha_m: \quad \frac{\sum_{k=0}^{m-1}\alpha_k x^k}{x^m} &= \sum_{k=0}^{m-1} \alpha^k \underbrace{\left(\frac{x^k}{x^m}\right)}_{\to 0} \overset{x\to+\infty}{\longrightarrow} 0. \end{aligned}$$

In particular we can conclude that

$$\forall \alpha_0,\dots,\alpha_m: \quad \sum_{k=0}^{m-1} \alpha^k x^k \neq x^m$$

because otherwise limit $\equiv 1$.

3. (a) Linear independence by 2.
Asssume $\exists \alpha_0,\dots,\alpha_n$ $(m := \max\{k : \alpha_k \neq 0\})$ not all zero with $\sum_{k=0}^m \alpha_k q_k = 0$

$$\begin{aligned} &\Rightarrow \quad \sum_{k=0}^{m} \alpha_k q_k = \sum_{k=0}^{m-1} \alpha_k q_k + \alpha_m q_m = 0 \\ &\Rightarrow \quad \sum_{k=0}^{m} \alpha_k x^k = (-\alpha_m) x^m \end{aligned}$$

contradiction to 2.

(b)

$$\begin{aligned} &\text{span}\{q_0,\dots,q_n\} = P_n(\mathbb{R}) \quad \text{by definition} \\ \Rightarrow \quad &\dim(P_n(\mathbb{R})) = n+1 \end{aligned}$$

4. Let $p = \sum_{k=0}^n \alpha_k q_k$ and $w = \sum_{k=0}^n \beta_k q_k$ and $\lambda \in \mathbb{R}$, then:

$$\begin{aligned} \mathrm{D}(\lambda p + w)(x) &= \left(\sum_{k=0}^n (\lambda\alpha_k + \beta_k) x^k\right)' = \sum_{k=0}^n (\lambda\alpha_k + \beta_k) k x^{k-1} \\ &= \lambda \sum_{k=0}^n \alpha_k k x^{k-1} + \sum_{k=0}^n \beta_k k x^{k-1} = \lambda \mathrm{D}(p)(x) + \mathrm{D}(q)(x) \end{aligned}$$

5. We have:

$$\begin{aligned} D: \; P_n(\mathbb{R}) \to P_n(\mathbb{R}), \; p = \sum_{k=0}^n \alpha_k q_k &\mapsto p' = \sum_{k=0}^{n-1} \alpha_{k+1}(k+1) q_k \\ \Phi: \; P_n(\mathbb{R}) \to \mathbb{R}^{n+1}, \; p = \sum_{k=0}^n \alpha_k q_k &\mapsto (\alpha_0,\dots,\alpha_n)^T \\ &= (\pi_0(p),\dots,\pi_n(p)) \end{aligned}$$

$\to$[Φ linear since π_j are linear (see lecture) and bijective since $\{q_0,\dots,q_n\}$ basis, $\Phi^{-1} : \mathbb{R}^{n+1} \to P_n(\mathbb{R})$, $(\alpha_0,\dots,,\alpha_n)^T \mapsto p = \sum \alpha_k q_k$]

Now consider: $F : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1},\ F(\alpha) := \Phi \circ D \circ \Phi^{-1}$

$$
\begin{array}{ccccc}
(\alpha_o, \dots, \alpha_n)^T & \begin{matrix}\mathbb{R}^{n+1}\\ \{e_1, \dots, e_{n+1}\}\end{matrix} & \xrightarrow{A} & \begin{matrix}\mathbb{R}^{n+1}\\ \{e_1, \dots, e_{n+1}\}\end{matrix} & (\alpha_1, 2\alpha_2, 3\alpha_3, \dots, n\alpha_n, 0) =: \beta \\
 & \downarrow \Phi^{-1} & & \Phi \uparrow & \\
p = \sum_{k=0}^{n} \alpha_k q_k & \begin{matrix}P_n(\mathbb{R})\\ \{q_1, \dots, q_{n+1}\}\end{matrix} & \overrightarrow{D} & \begin{matrix}P_n(\mathbb{R})\\ \{q_1, \dots, q_{n+1}\}\end{matrix} & p' = \sum_{k=0}^{n} \beta_k q_k
\end{array}
$$

$$
\begin{aligned}
F(\alpha) &= (\Phi \circ D \circ \Phi^{-1})(\alpha) \\
&= (\Phi \circ D)\left(\sum_{k=0}^{n} \alpha_k q_k\right) \\
&= \Phi\left(\sum_{k=0}^{n-1} \alpha_{k+1}(k+1) q_k\right) \\
&= (\alpha_1, 2\alpha_2, 3\alpha_3, \dots, n\alpha_n, 0)^T
\end{aligned}
$$

To obtain the matrix representation we have to evaluate F on the standard basis $\{e_1, \dots, e_{n+1}\}$:

$$
A = \begin{pmatrix} | & & | \\ F(e_1) & \dots & F(e_{n+1}) \\ | & & | \end{pmatrix} = \begin{pmatrix}
0 & 1 & 0 & 0 & \cdots & 0 \\
0 & 0 & 2 & 0 & \cdots & 0 \\
\vdots & \vdots & 0 & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \ddots & n-1 & 0 \\
0 & 0 & 0 & \cdots & 0 & n \\
0 & 0 & 0 & \cdots & 0 & 0
\end{pmatrix}
$$

PROGRAMMING

---

5 Appetizer: Gradient, Steepest Descent and Conjugate Gradient Method

Background

Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite (spd). Then A is in particular invertible, so that the linear system $Ax = b$ has a unique solution $x^* \in \mathbb{R}^n$ for all $b \in \mathbb{R}^n$. Let us relate this linear system to an optimization problem. For this purpose we define for a fixed spd matrix A and fixed right-hand side b the function

$$f := f_{A,b} : \mathbb{R}^n \to \mathbb{R}, x \to \tfrac{1}{2} x^T A x - b^T x.$$

Then one can show the equivalence

$$Ax^* = b \quad \iff \quad x^* = \arg\min_{x \in \mathbb{R}^n} f(x).$$

In words, x^* solves the linear system on the left-hand side if and only if x^* is the unique minimizer of the functional f. In fact, you will learn in the next semester that the condition $Ax^* = b$ is the necessary first-order optimality condition:

$$0 = \nabla f(x) = Ax - b.$$

Due to the convexity of f this condition is also sufficient. Consequently, solving linear systems which involve spd matrices is equivalent to solving the associated optimization problem above, i.e., minimizing the function $f(x) = \frac{1}{2} x^T A x - b^T x$. Thus, in this context iterative methods for linear systems, such as the Richardson iteration, can also be interpreted as optimization algorithms. Let us consider the (relaxed) Richardson iteration for $Ax = b$, i.e., $x_{k+1} = (I - \theta A)x_k + \theta b$. After some minor manipulations and making use of $\nabla f(x_k) = Ax_k - b$ we arrive at the equivalent formulation

$$x_{k+1} = x_k - \theta \nabla f(x_k).$$

The latter is what is called a gradient method. A step from x_k into (an appropriately scaled) direction of the gradient $\nabla f(x_k)$ yields a decrease in the objective function f , i.e., $f(x_{k+1}) \leq f(x_k)$. Along the Richardson/ Gradient method the scaling (also called step size) θ is fixed. However, one could also choose a different θ_k in each iteration step. This gives the more general version

$$x_{k+1} = x_k - \theta_k \nabla f(x_k). \tag{1}$$

The well known method of steepest descent is given by choosing

$$\theta_k = \frac{r_k^\top r_k}{r_k^\top A r_k}, \tag{2}$$

where $r_k := Ax_k - b$ is the k-th residual. This choice can be shown to be optimal in terms of convergence speed. Even general, one can think of using a different preconditioner N_k in each iteration step – this will later correspond to Newton-type optimization algorithms.

Task

Consider the following setting:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 10 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad x_0 = \begin{bmatrix} 4 \\ 1.4 \end{bmatrix}.$$

Convince yourself that A is spd. Determine the minimal and maximal eigenvalue of A, i.e., λ_{min} and λ_{max}, respectively. What is the solution to $Ax = b$? Now extend your code (in particular `iter_solve()`) from previous sheets:

1. Implement the **steepest descent method** (1) by choosing the stepsize θ_k from (2) in each iteration step.
2. Find a way to apply the **conjugate gradient method** to solve a system $Ax = b$, where A is spd.
   Hint: You can either implement it on your own or find a SciPy routine (for the latter: you can collect all iterates x_k by using the callback interface).
3. **Test:** Solve the above problem with the following routines:
   (a) Richardson method with $\theta = \frac{2}{\lambda_{\text{max}}}$
   (b) Richardson method with $\theta = 0.9 \cdot \frac{2}{\lambda_{\text{max}}}$
   (c) Richardson method with optimal $\theta = \frac{2}{\lambda_{\text{min}} + \lambda_{\text{max}}}$
   (d) Steepest descent method
   (e) conjugate gradient method

   Generate the following two plots:

   1. Plot the iterates x_k for all the runs into the same 2d plot (use different colors).
   2. Plot the function values $f(x_k) = \frac{1}{2} x_k^T A x_k - b^T x_k$ for each iterate and all runs into a second plot (use different colors).

(10 Points)

Solution:

```python
# -*- coding: utf-8 -*-
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.patches import Ellipse
import importlib

# import iterative solvers from previous sheets
iterSolver = importlib.import_module("prog-LinearIterations_JacRich_solution")


if __name__ == "__main__":
    # ---------------
    # PROBLEM SETUP
    # ---------------
    eigmin, eigmax = 2, 10
    A = np.array([[eigmin, 0], [0, eigmax]])
    b = np.array([0, 0])
    x = np.array([0, 0])
    x0 = np.array([4, 1.4])

    theta_max = 2./eigmax
    theta_bad = 0.99 * theta_max
    theta_opt = 2 / (eigmin+eigmax)

    # ---------------
    # METHOD SETUP
    # ---------------
    maxiter = 200
    legend = ["Richardson max", "Richardson bad", "Richardson opt",
              "Steepest Descent", "Conjugate Gradient"]
    methods = ["Richardson", "Richardson", "Richardson",
               "steepestDescent", "CG"]
    theta = [theta_max, theta_bad,  theta_opt,
             -1, -1]
    colors = ["y", "c", "b",
              "g", "r"]
    numberRuns = len(methods)

    # ---------------
    # SOLVE
    # ---------------
    X = []
    F = []
    for i, method in enumerate(methods):
        _X = iterSolver.main(A, b, x0, maxiter,
                                {method: theta[i]}, plot=0, verbose=0)[0]
        X += [np.array(_X)]
        F += [[0.5 * np.dot(x, A.dot(x)) for x in _X]]

    # ------------- #
    # plot iterates
    # ------------- #
    fig, ax = plt.subplots()
    for i, _X in enumerate(X):
        plt.plot(_X[:, 0], _X[:, 1], colors[i]+"o-")
    # add level sets
    m = 200
    for k in range(m):
        e = Ellipse(xy=np.zeros(2), width=eigmax*0.8**k,
                    height=eigmin*0.8**k, angle=0, fill=False)
```

```python
        ax.add_artist(e)
    plt.legend(legend)
    plt.axis('equal')
    plt.show()
    # -------------- #
    # plot objective
    # -------------- #
    plt.figure("objective-function")
    for i, f in enumerate(F):
        plt.plot(f, colors[i]+"x-")
    plt.legend(legend)
    plt.show()
```

Total Number of Points = 32 (T:22, P:10)