SOLUTION 12 – Advanced Numerical Analysis

Prof. D. Kressner
M. Steinlechner

1 ▶ Rosenbrock function with equality constraint

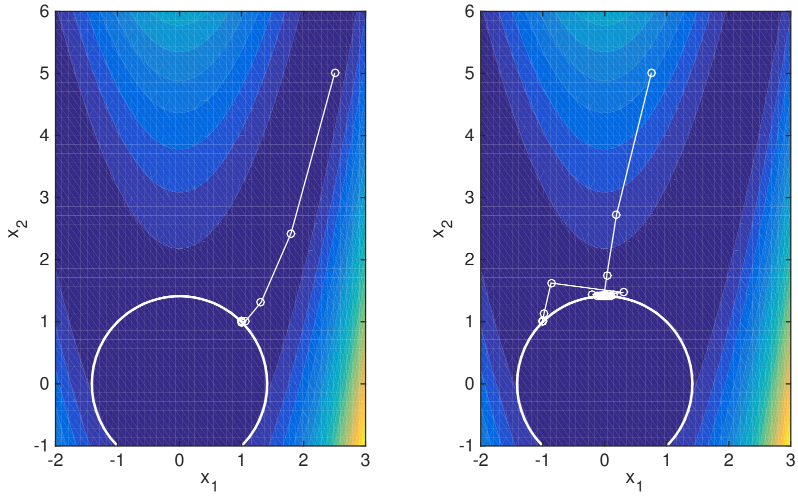
In this exercise, we want to find the minimum of the Rosenbrock function subject to an equality constraint:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^2} \quad & f(\mathbf{x}) := 100\big(x_2 - x_1^2\big)^2 + (1-x_1)^2 \\ \text{subject to:} \quad & x_1^2 + x_2^2 - 2 = 0. \end{aligned}$$

(a) Write down the KKT conditions for this system, equivalent to finding the zeros of a nonlinear function. Apply the Newton method to derive equation (5.34) in the script for this problem. Finally, substitute μ^{k+1} to arrive at the linear system (5.36) corresponding to the local quadratic program.

(b) Find the solution of the constraint optimization problem by a simple local SQP algorithm: Beginning from the initial guess $(x_1^0, x_2^0, \mu^0) = (2.5, 5, 1)$, solve the quadratic program around the current iterate (x_1^k, x_2^k, μ^k) and update: $\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{p}^k$. Compare to the initial guess $(x_1^0, x_2^0, \mu^0) = (0.75, 5, 1)$. What do you observe?

```
function [x,mu] = rosenbrock()
    % Cost and constraint functions and their derivatives
    f = @(x,y) 100*(y-x.^2).^2 + (1-x).^2;
    h = @(x) x(1)^2 + x(2)^2 - 2;
    dF = @(x) [400*x(1)^3-400*x(1)*x(2)-2+2*x(1); -200*(x(1)^2-x(2))];
    H = @(x,mu) [1200*x(1)^2+2-400*x(2)+2*mu, -400*x(1); ...
                 -400*x(1),                   200+2*mu];
    dh = @(x) [2*x(1); 2*x(2)];

    % Plot function and constraint
    [X,Y] = meshgrid( linspace(-2,3,50), linspace(-1,6,50));
    contourf(X,Y,f(X,Y),20,'linestyle','none');
    hold on
    contour( X,Y,X.^2 + Y.^2, [2,2],'w','linewidth',2)

    % solver for local quadratic program
    function [p, mu] = solveSQP( x, mu )
        A = [ H(x,mu), -dh(x);...
              dh(x)' ,     0 ];
        b = [ -dF(x); -h(x)];

        y = A \ b;
        p = y(1:2);
        mu = y(3);
    end

    x = [2.5;5]; % first initial guess
    %x = [0.75;5]; % second initial guess

    mu = 1;
    for i = 1:100
        [p, mu_new] = solveSQP( x(:,end), mu(:,end) );
        x = [x, x(:,end)+p];
        mu = [mu, mu_new];
    end

    plot( x(1,:), x(2,:), '-ow' )
    xlabel('x_1')
    ylabel('x_2')
    set(gca,'fontsize',14)
    axis equal
end
```



(a) Initial guess $(2.5, 5, 1)$ (b) Initial guess $(0.75, 5, 1)$

2 ▶ Subdifferential

In this exercise, we want to explicitely calculate the subdifferential, the set of all subgradients, of simple functions.

a) $f(x) = |x|$ at $x_0 = 1$ and $x_0 = 0$.

This function is differentiable at all points $x \neq 0$ *and* $f'(x) = \text{sign}(x)$. *Therefore, the subdifferential at points* $x_0 = 1$ *is given by* $\partial f(-1) = \text{sign}(1) = 1$. *At point* $x_0 = 0$, *we have from the definition of the subgradient* g *that it has to hold*

$$|x| \equiv f(x) \geq f(x_0) + \langle g, x - x_0 \rangle = gx$$

which is true for all g *with* $|g| \leq 1$. *Hence, we obtain* $\partial f(0) = [-1, 1]$.

b) $f(x) = \max\left\{0, \frac{1}{2}(x^2 - 1)\right\}$ at $x_0 = \pm 1$.

f *is differentiable everywhere except at* $x_0 = \pm 1$. *At these two points, the subdifferential is given by* $\partial f(-1) = [-1, 0]$ *and* $\partial f(1) = [0, 1]$.

c) $f(x) = -\sqrt{x}$ with $x \in \{y \in \mathbb{R} \,|\, y \geq 0\}$ at $x_0 = 0$.

This function is not subdifferentiable at $x_0 = 0$ on its domain $\text{dom}(f) = [0, \infty[$.

3 ▶ Proximal operators

a) Consider the ℓ_1-norm of a vector, $g(x) = \|x\|_1 = \sum_{i=1}^{n} |x_i|$ for $x \in \mathbb{R}^n$ and let $\lambda > 0$ be a parameter. Show that we can compute the proximal operator for this function explicitly in the following form (for the ith component of $\text{prox}_{\lambda g}(x)$):

$$\left(\text{prox}_{\lambda g}(x)\right)_i = \begin{cases} x_i - \lambda & \text{if } x_i > \lambda, \\ x_i + \lambda & \text{if } x_i < -\lambda, \\ 0 & \text{if } -\lambda \leq x_i \leq \lambda. \end{cases},$$

which can be written in short form as $\left(\text{prox}_{\lambda g}(x)\right)_i = \max\{|x_i| - \lambda, 0\} \,\text{sign}(x_i)$.

This operator is also known as the *shrinkage* (soft thresholding) operator

$$S_\lambda(x) := \text{prox}_{\lambda g}(x).$$

We have that the proximal operator of $g(x) = \|x\|_1$ *is given by:*

$$\text{prox}_{\lambda g}(z) = \underset{y \in \mathbb{R}^p}{\text{argmin}}\, \lambda \|y\|_1 + \frac{1}{2}\|y - z\|_2^2.$$

Taking the derivative w.r.t. the i-th coordinate and setting it to zero:

$$0 \stackrel{!}{=} \frac{\partial}{\partial y_i}\left(\lambda\|y\|_1 + \frac{1}{2}\|y-z\|_2^2\right) = \frac{\partial}{\partial y_i}\left(\lambda|y_i| + \frac{1}{2}(y_i - z_i)^2\right)$$

as the other terms to not depend on y_i. Note that the absolute value function is not differentiable in 0. Hence, we look at the cases $y_i < 0$ and $y_i > 0$ separately: If $y_i > 0$, then we obtain

$$0 = \lambda + (y_i - z_i) \quad \Rightarrow \quad y_i = z_i - \lambda.$$

This can only hold if $z_i > \lambda$ (otherwise $y_i < 0$, contrary to the assumption). On the other hand, if $y_i < 0$, then we obtain

$$0 = -\lambda + (y_i - z_i) \quad \Rightarrow \quad y_i = z_i + \lambda.$$

This can only hold if $z_i < -\lambda$ (otherwise $y_i > 0$, contrary to the assumption). For remaining case $-\lambda \leq z_i \leq \lambda$, $y_i = 0$.
Hence we have in total:

$$(\text{prox}_{\lambda g}(z))_i = \begin{cases} z_i - \lambda & \textit{for } z_i > \lambda \\ z_i + \lambda & \textit{for } z_i < -\lambda \\ 0 & \textit{for } -\lambda \leq z_i \leq \lambda \end{cases}$$

which can also be written in short form as

$$(\text{prox}_{\lambda g}(z))_i = \max(|z_i| - \lambda, 0)\,\text{sign}(z_i),$$

or, for all components of the vector at once,

$$\text{prox}_{\lambda g}(z) = \max(|z| - \lambda, 0) \otimes \text{sign}(z).$$

*b) Let A be an $n \times n$ matrix and $g(A) := \sigma_1(A) + \cdots + \sigma_n(A)$, where σ_j is the jth singular value of A. What is $\text{prox}_{\lambda g}(A)$?

Note: In the definition of $\text{prox}_{\lambda g}(A)$, the 2-norm of the vector needs to be replaced by the Frobenius norm of the matrix.

We have

$$\text{prox}_{\lambda g}(A) = \underset{X}{\text{argmin}}\left(\lambda g(A) + \frac{1}{2}\|X - A\|_F^2\right).$$

Furthermore, let $A = USV^T$ be the singular value decomposition of A. Then it can be shown that

$$\text{prox}_{\lambda g}(A) = U\,\text{diag}(\rho_1, \ldots, \rho_n)V^T,$$

where

$$\rho_i = \begin{cases} \sigma_i - \lambda, & \text{if } \sigma_i \geq \lambda, \\ 0, & \text{if } -\lambda \leq \sigma_i \leq \lambda, \\ \sigma_i + \lambda, & \text{if } \sigma_i \leq -\lambda. \end{cases}$$

are the thresholded singular values.