MA40050: Numerical Optimisation & Large–Scale Systems

Model Solutions to Problem Sheet 3

---

1. Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex and let x_* be a local minimiser of f.

   Assume, for contradiction, that x_* is **not** a global minimiser. Then

   $$\exists z \in \mathbb{R}^N : f(z) < f(x_*).$$

   Let $x = \alpha x_* + (1-\alpha)z$. Since f is convex,

   $$\begin{aligned} f(x_*) &= \alpha f(x_*) + (1-\alpha) f(x_*) \\ &> \alpha f(x_*) + (1-\alpha) f(z) \\ &\geq f(\alpha x_* + (1-\alpha)z) = f(x), \quad \forall \alpha \in [0,1] \end{aligned}$$

   and so x_* is not a local minimiser of f which is a contradiction. Hence, x_* is a global minimiser of f.

2. (a) Clearly the result holds for $k = 0$. Let us assume $x_{2k} = (0, 1 - 5^{-k})^T$.
   Since $f(x) = (x_1 - x_2)^2 + 2(x_1 - x_2) + x_1^2$,

   $$\nabla f(x) = \begin{pmatrix} 4x_1 - 2(x_2 - 1) \\ -2x_1 + 2(x_2 - 1) \end{pmatrix}$$

   and so $\nabla f(x_{2k}) = 5^{-k}(2, -2)^T$. Thus, the direction of steepest descent is $s_{2k} = (-1, 1)^T$ (only the direction of s_k matters), and so

   $$x_{2k+1} = x_{2k} + \alpha s_{2k} = \begin{pmatrix} -\alpha \\ \alpha + 1 - 5^{-k} \end{pmatrix}$$

   where α is chosen such that it minimises

   $$\phi(\alpha) := f(x_{2k+1}) = (-2\alpha - 1 + 5^{-k})^2 + 2(-2\alpha - 1 + 5^{-k}) + \alpha^2.$$

   We have $\phi'(\alpha) = -4(-2\alpha - 1 + 5^{-k}) - 4 + 2\alpha = 10\alpha - 4/5^k$ and $\phi''(\alpha) = 10$. Hence, the unique critical point of ϕ is $\alpha_* = 2/5^{k+1}$ and it is a minimum. Thus,

   $$x_{2k+1} = \begin{pmatrix} -2/5^{k+1} \\ 1 - 3/5^{k+1} \end{pmatrix}. \tag{1}$$

   Similarly,

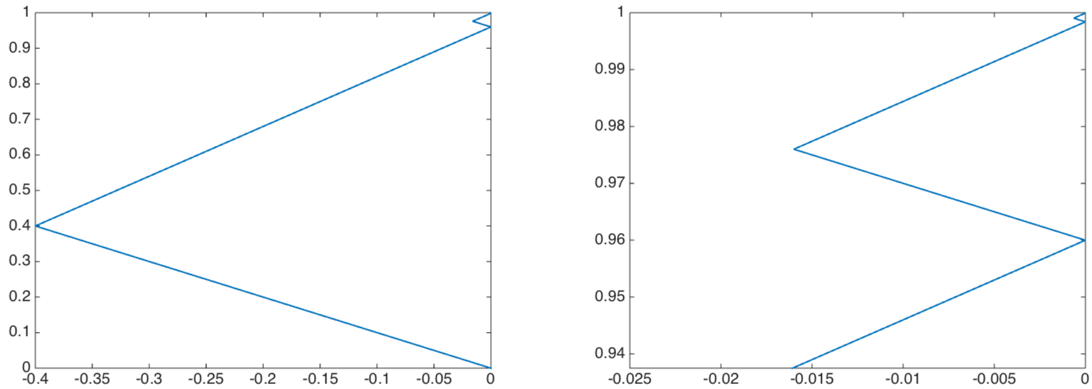
   $$\nabla f(x_{2k+1}) = 5^{-(k+1)} \begin{pmatrix} -8 + 6 \\ 4 - 6 \end{pmatrix}, \quad s_{2k+1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad x_{2k+2} = \begin{pmatrix} \alpha - 2/5^{k+1} \\ \alpha + 1 - 3/5^{k+1} \end{pmatrix}.$$

Here,

$$\phi(\alpha) := f(x_{2k+2}) = (-1+5^{-(k+1)})^2 + 2(-1+5^{-(k+1)}) + (-2/5^{k+1}+\alpha)^2\,,$$

$\phi'(\alpha) = 2(-2/5^{k+1}+\alpha)$ and $\phi''(\alpha) = 2$. Hence, the unique minimum is $\alpha_* = 2/5^{k+1}$ and $x_{2k+2} = (0, 1-5^{-(k+1)})^T$ which completes the induction step.

The sequence clearly converges to the local minimum $x_* = (0,1)^T$. Here is a sketch of the iterates:



(b) Since

$$|x_{2k} - x_*| = 5^{-k} \quad \text{and} \quad |x_{2k+1} - x_*| = \sqrt{13}/5^{k+1}$$

we have

$$\frac{|x_{2k+2} - x_*|}{|x_{2k} - x_*|} = \frac{1}{5} \quad \text{and} \quad \frac{|x_{2k+3} - x_*|}{|x_{2k+1} - x_*|} = \frac{1}{5}$$

and we can choose $\xi_k = C5^{-k/2}$. To find a suitable C, let us pick the smallest value of C such that

$$1 = |x_0 - x_*| \le \xi_0 = C \quad \text{and} \quad \sqrt{13}/5 = |x_1 - x_*| \le \xi_1 = C5^{-1/2}\,.$$

This implies $C = \sqrt{13/5}$ and so $\xi_k = \sqrt{13/5^{k+1}}$ which converges q-linearly to 0 with q-factor $5^{-1/2}$. Hence, $x_k \to x_*$ r-linearly with r-factor $5^{-1/2}$.

3. Model code is available on the course website.

As predicted Algorithm 4.2 converges extremely poorly, especially for the more difficult starting point $x_0 = (-1.2, 1)^T$. More than 10000 iterations are necessary to achieve a tolerance of 10^{-10} for $\theta_{sd} = 10^{-3}$ (for both starting points) and $\sim$900 iterations for $\theta_{sd} = 0.37$ for the easier starting point $x_0 = (1.2, 1.2)^T$, which is close to the solution and $\sim$2000 iterations for the more difficult starting point $x_0 = (-1.2, 1)^T$ which is further away from the exact solution.

4. (a) First note that $\nabla f(x_*) = DR(x_*)^T R(x_*) = 0$. Furthermore, since

$$\nabla^2 f(x_*) = DR(x_*)^T DR(x_*) + \sum_{j=1}^{N} \underbrace{R_j(x_*)}_{=0} \nabla^2 R_j(x_*) = DR(x_*)^T DR(x_*)$$

and $DR(x_*)$ was assumed to be of full rank, we have

$$h^T \nabla^2 f(x_*) h = (DR(x_*)h)^T \underbrace{(DR(x_*)h)}_{=:\, y \neq 0} = y^T y > 0, \quad \text{for all } h \neq 0,$$

and so $\nabla^2 f(x_*) > 0$ and x_* is a strict local minimiser of f.

(b) Since $\nabla^2 f(x_*) = DR(x_*)^T DR(x_*) > 0$, it follows as usual from Lemma 2.1 that

$$\exists R > 0 : \forall x_n \in \overline{B}_R(x_*) : \quad DR(x_n)^T DR(x_n) \text{ invertible and } x_{n+1} \text{ well defined.}$$

Suppose $x_n \in \overline{B}_R(x_*)$. Since $R(x_*) = 0$,

$$\begin{aligned} x_{n+1} - x_* &= x_n - x_* - \left(DR(x_n)^T DR(x_n)\right)^{-1} DR(x_n)^T R(x_n) \\ &= \left(DR(x_n)^T DR(x_n)\right)^{-1} DR(x_n)^T \Big(R(x_*) - R(x_n) - DR(x_n)(x_* - x_n)\Big) \end{aligned} \tag{2}$$

As in the Proof of Theorem 3.2, we can use the IMVT (Theorem 2.5) to show that

$$R(x_*) - R(x_n) - DR(x_n)(x_* - x_n) = \left[\int_0^1 \left(DR(x_n + t(x_* - x_n)) - DR(x_n)\right) \mathrm{d}t\right] (x_* - x_n)$$

and hence (using the Lipschitz continuity of DR near x_* with constant $L > 0$)

$$\begin{aligned} \|R(x_*) - R(x_n) - DR(x_n)(x_* - x_n)\| &\leq \int_0^1 \|DR(x_n + t(x_* - x_n)) - DR(x_n)\| \,\mathrm{d}t\, |x_n - x_*| \\ &\leq L \int_0^1 |t - 1| \,\mathrm{d}t\, |x_n - x_*|^2 = \frac{L}{2} |x_n - x_*|^2 . \end{aligned}$$

Using this bound together with (2), we get

$$|x_{n+1} - x_*| \leq \frac{L}{2} \left\| \left(DR(x_n)^T DR(x_n)\right)^{-1} \right\| \left\| DR(x_n) \right\| |x_n - x_*|^2 \leq C |x_n - x_*|^2 , \tag{3}$$

where the constant C depends on L, on $\max_{x \in \overline{B}_R(x_*)} \|DR(x)\|$, and – again through Lemma 2.1 – on $\|(DR(x_*)^T DR(x_*))^{-1}\|$.
Now, by choosing $x_0 \in \overline{B}_r(x_*)$ with $r = \min(R, \frac{1}{2C})$, we have

$$|x_{n+1} - x_*| \leq \tfrac{1}{2} |x_n - x_*| \leq \ldots \leq \left(\tfrac{1}{2}\right)^{n+1} r$$

and it follows as in the proof of Theorem 3.2 by induction that $x_n \to x_*$. The q-quadratic convergence follows from (3).