ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 7

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Central path

a) First we reformulate the problem to be in the form seen in the lecture:

$$\begin{array}{lll} \min & y_1 + 2y_2 & \\ \text{s.t.} & y_1 + y_2 & \geq 4 \\ & y_2 & \geq 0 \end{array}$$

Then we identify:

$$A^T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad c = \begin{pmatrix} 4 \\ 0 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Define slack variables s_1, s_2:

$$\begin{array}{lll} \max & -y_1 - 2y_2 & \\ \text{s.t.} & y_1 + y_2 - s_1 & = 4 \\ & y_2 - s_2 & = 0 \\ & s_1,\ s_2 & \geq 0 \end{array} ,$$

which yields:

$$\begin{aligned} s_1 &= y_1 + y_2 - 4 \\ s_2 &= y_2 \end{aligned}$$

b) The logarithmic barrier function for this problem is given by:

$$\begin{aligned} \phi(y) &= -\sum_{j=1}^{n} \log(s_j) \\ &= -\log(y_1 + y_2 - 4) - \log(y_2). \end{aligned}$$

c) Since the central path $y(\mu)$ is the minimizer of $b^\top y + \frac{1}{\mu}\phi(y)$, we compute the gradient of $\phi(y)$:

$$\begin{aligned} \frac{\partial}{\partial y_1}\,\phi(y) &= -\frac{1}{y_1 + y_2 - 4} \\ \frac{\partial}{\partial y_2}\,\phi(y) &= -\frac{1}{y_1 + y_2 - 4} - \frac{1}{y_2} \end{aligned}$$

or directly:

$$\begin{aligned} \nabla\phi(y) &= -A \begin{pmatrix} \frac{1}{s_1} \\ \frac{1}{s_2} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{y_1+y_2-4} \\ -\frac{1}{y_1+y_2-4} - \frac{1}{y_2} \end{pmatrix} \end{aligned}$$

Thus,

$$\begin{aligned}
\nabla\left\{b^\top y+\frac{1}{\mu}\phi(y)\right\} &= b-\frac{1}{\mu}A\begin{pmatrix}\frac{1}{s_1}\\ \frac{1}{s_2}\end{pmatrix}\\
&= \begin{pmatrix}1-\frac{1}{\mu(y_1+y_2-4)}\\ \\ 2-\frac{1}{\mu(y_1+y_2-4)}-\frac{1}{\mu y_2}\end{pmatrix}
\end{aligned}$$

To obtain the minimum of $\phi(y)$ we set the gradient to zero and solve for y_1, y_2. This yields:

$$\begin{aligned}
1 =&\frac{1}{\mu(y_1+y_2-4)}\\
\mu(y_1+y_2-4) =&1\\
y_1+y_2-4 =&\frac{1}{\mu}
\end{aligned}$$

as well as

$$\begin{aligned}
\frac{1}{\mu}\left(\frac{1}{y_1+y_2-4}+\frac{1}{y_2}\right) =&2\\
\left(\frac{1}{y_1+y_2-4}+\frac{1}{y_2}\right) =&2\mu\\
\left(-\mu+\frac{1}{y_2}\right) =&2\mu\\
-y_2\mu+1 =&2y_2\mu\\
1 =&y_2\mu\\
y_2 =& \frac{1}{\mu}
\end{aligned}$$

leading to

$$y_1+\frac{1}{\mu}-4=\frac{1}{\mu}$$

and we arrive at

$$y_1=4 \qquad y_2=\frac{1}{\mu}.$$

Hence the analytic path $y(\mu)$ is given by:

$$y(\mu)=\begin{pmatrix}4\\ \frac{1}{\mu}\end{pmatrix}$$

Letting $\mu\to\infty$ we obtain an optimal solution as $\binom{4}{0}$.

Exercise 2: Strict vs. Strong Convexity

i) $f:[0,1]\to\mathbb{R}$ defined by $f(0)=f(1)=1$ and $f(x)=0$ for $x\in(0,1)$ is convex, but not continuous.

ii) $f(x)=|x|$ is convex, but not strictly convex. In fact, note that $f(\lambda x)=|\lambda|f(x)$, so if $y=0$ then $f(\lambda x+(1-\lambda)y)=f(\lambda x)\not<\lambda f(x)=\lambda f(x)+(1-\lambda)f(y)$.

iii) $f(x)=e^x$ is strictly convex, but not strongly convex. In fact, $f''(x)=e^x>0\ \forall x\in\mathrm{dom}(f)$, so it is strictly convex. It is not strongly convex because its second derivative can be arbitrarily close to zero.

iv) $f(x) = x^2$ is strongly convex with modulus 2 because $f''(x) = 2 \; \forall x \in \mathrm{dom}(f)$.

Exercise 3: Convergence of the Newton method
Note that for $k \geq 1$, $\|\nabla f(x_t)\|_2^k \leq \left(\frac{L}{2\sigma^2}\right)^k \|\nabla f(x_{t-1})\|_2^{2k}$. Thus

$$\frac{L}{2\sigma^2}\|\nabla f(x_T)\|_2 \leq \frac{L}{2\sigma^2}\left(\frac{L}{2\sigma^2}\|\nabla f(x_{T-1})\|_2^2\right) \leq \frac{L}{2\sigma^2}\frac{L}{2\sigma^2}\left(\frac{L}{2\sigma^2}\frac{L}{2\sigma^2}\|\nabla f(x_{T-2})\|_2^4\right) \leq \cdots \leq \left(\frac{L}{2\sigma^2}\right)^{2^T} \underbrace{\|\nabla f(x_0)\|}_{\leq \frac{\sigma^2}{L}}{}^{2^T},$$

which in turn fulfills

$$\left(\frac{L}{2\sigma^2}\right)^{2^T} \|\nabla f(x_0)\|_2^{2^T} \leq \left(\frac{1}{2}\right)^{2^T}.$$

Exercise 4: 'Invariance' of the Newton step under linear transformations
By applying the chain rule, we obtain

$$\begin{array}{rcl} \nabla \tilde{f}(x) & = & A^T \nabla f(Ax) \\ \nabla^2 \tilde{f}(x) & = & A^T \nabla^2 f(Ax) A \end{array}$$

Hence we get

$$\begin{array}{rcl} x^{k+1} & \overset{(\mathrm{def})}{=} & x^k - (\nabla^2 \tilde{f}(x^k))^{-1} \cdot \nabla \tilde{f}(x^k) \\ & = & x^k - A^{-1}(\nabla^2 f(Ax))^{-1}(A^T)^{-1} \cdot A^T \nabla f(Ax) \\ & = & x^k - A^{-1}(\nabla^2 f(Ax))^{-1} \nabla f(Ax) \end{array}$$

and

$$\begin{array}{rcl} Ax^{k+1} & = & Ax^k - AA^{-1}(\nabla^2 f(Ax))^{-1} \nabla f(Ax) \\ & = & Ax^k - (\nabla^2 f(Ax^k))^{-1} \nabla f(Ax^k) \\ & \overset{(Ax^k = y^k)}{=} & y^k - (\nabla^2 f(y^k))^{-1} \nabla f(y^k) \\ & \overset{(\mathrm{def})}{=} & y^{k+1} \end{array}$$