Solutions to exercises for Part 1.

1(a). The first-order optimality conditions are that there exist vectors of Lagrange multipliers $y_{\mathcal{E}*}$ and $y_{\mathcal{I}*}$ such that

$$\begin{array}{rll} c_{\mathcal{E}}(x_*) = 0 & \text{and } \; c_{\mathcal{I}}(x_*) \geq 0 & \text{(primal feasibility),} \\ g(x_*) - A_{\mathcal{E}}^T(x_*) y_{\mathcal{E}*} - A_{\mathcal{I}}^T(x_*) y_{\mathcal{I}*} = 0 & \text{and } \; y_{\mathcal{I}*} \geq 0 & \text{(dual feasibility) and} \\ c_i(x_*)[y_*]_i = 0 & \text{for all } \; i \in \mathcal{I} & \text{(complementary slackness).} \end{array}$$

1(b). The second-order optimality conditions are that necessarily

$$s^T H(x_*, y_*) s \geq 0 \; \text{ for all } \; s \in \mathcal{N}_+,$$

where

$$\mathcal{N}_+ = \left\{ s \in \Re^n \;\middle|\; \begin{array}{l} s^T a_i(x_*) = 0 \text{ if } \; i \in \mathcal{E} \\ s^T a_i(x_*) = 0 \text{ if } \; i \in \mathcal{I} \; \text{ \& both } \; c_i(x_*) = 0 \text{ \& } [y_*]_i > 0 \text{ and} \\ s^T a_i(x_*) \geq 0 \text{ if } \; i \in \mathcal{I} \; \text{ \& both } \; c_i(x_*) = 0 \text{ \& } [y_*]_i = 0 \end{array} \right\},$$

and $y_* = (y_{\mathcal{E}*}^T, \quad y_{\mathcal{I}*}^T)^T$.

2(a). The problem might be non-differentiable because small perturbations in x may cause different terms $f_i(x)$ to define the objective $f(x)$. For example, suppose $m = 2$, $f_1(x) = x + 1$ and $f_2(x) = -x + 1$. Then for $x \geq 0$, $f(x) = x + 1$ while for $x \leq 0$, $f(x) = -x + 1$, and there is a derivative discontinuity at $x = 0$. It might also be non-differentiable because of the $|\cdot|$ term. For instance if $m = 1$ and $f_1(x) = x$, $f(x)$ is non-differentiable at $x = 0$.

2(b). Clearly $|f_i(x)| \leq u$ is equivalent to $-u \leq f_i(x) \leq u$. Minimizing the largest $|f_i(x)|$ is equivalent to minimizing the largest upper bound on $|f_i(x)|$.

2(c). The constraints $-u \leq f_i(x) \leq u$ may be rewritten as $f_i(x) + u \geq 0$ and $u - f_i(x) \geq 0$. Let y_i^{L} and y_i^{U} (respectively) be Lagrange multipliers for these constraints, and let $A(x)$ be the Jacobian of the vector of f_i.

First-order necessary optimality conditions are that the y^{L} and y^{U} satisfy

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} A(x) \\ e^T \end{pmatrix} y^{\text{L}} - \begin{pmatrix} -A(x) \\ e^T \end{pmatrix} y^{\text{U}} = 0$$

and that

$$(f^{\max} + f_i(x)) y_i^{\text{L}} = 0 \; \text{ and } \; (f^{\max} - f_i(x)) y_i^{\text{U}} = 0,$$

where $f^{\max}$ is the optimal objective value. This is to say that

$$\begin{array}{rl} A(x)(y^{\text{L}} - y^{\text{U}}) & = 0 \\ e^T(y^{\text{L}} + y^{\text{U}}) & = 1 \; \text{ and } \; (y^{\text{L}}, y^{\text{U}}) \geq 0 \end{array}.$$

If $f^{\max} > 0$ only one of the pair $(y_i^{\text{L}}, y_u^{\text{L}})$ can be nonzero.

Solutions to exercises for Part 2.

1(a). The gradient of the objective function is $g = Hx$ and $g(x_*) = Hx_* = H0 = 0$, so that x_* is a stationary point which is a minimum, since H is positive definite.

1(b). Line-search in direction p from x gives

$$\begin{aligned} f(x+\alpha p) &= \tfrac{1}{2}(x+\alpha p)^T H (x+\alpha p) \\ &= \tfrac{1}{2}\alpha^2 p^T H p + \alpha p^T H x + \tfrac{1}{2} x^T H x. \end{aligned}$$

Hence, the exact line-search condition $\dfrac{df}{d\alpha} = 0$, using $g = g(x) = Hx$ is equivalent to

$$\alpha p^T H p + p^T g = 0 \;\Leftrightarrow\; \alpha = -\frac{p^T g}{p^T H p},$$

where we have used the positive definiteness of H, which ensures that $p^T H p > 0$ for all $p \neq 0$.

1(c). If x_1 is chosen as in the question, then the gradient

$$g_1 \;=\; (\sigma, 0, \ldots, 0, 1)^T \;=\; -p_1$$

is the steepest descent direction. Next, compute

$$-p_1^T g_1 = \sigma^2 + 1 = 2 \;\text{ and }\; p_1^T H p_1 = \lambda_1 + \lambda_n,$$

and using the step-length from (b), it follows that

$$\alpha_1 = \frac{2}{\lambda_1 + \lambda_n}.$$

Now compute the next iterate as

$$x_2 = x_1 + \alpha_1 p_1 = \begin{pmatrix} \frac{\sigma}{\lambda_1} \\ 0 \\ \vdots \\ 0 \\ \frac{1}{\lambda_n} \end{pmatrix} + \frac{2}{\lambda_1 + \lambda_n} \begin{pmatrix} -\sigma \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix} = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \begin{pmatrix} \frac{\sigma}{\lambda_1} \\ 0 \\ \vdots \\ 0 \\ \frac{1}{\lambda_n} \end{pmatrix}.$$

Each subsequent iteration only differs from iteration 1 by the factor $\dfrac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$. Note that the step-length is independent of this factor. Each iteration "adds" one factor to the expression for x_{k+1} giving the desired formula.
1(c) *(i)*. If $\lambda_1 = \lambda_n$, then $x_2 = 0$ is optimal.

1(c) *(ii)*. If $\lambda_1 \gg \lambda_n$, then steepest descent converges very slowly, since $\lambda_1 - \lambda_n \simeq \lambda_1$, the sequence $\dfrac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$ approaches zero very slowly. The rate of convergence is linear, since

$$\frac{\|x_{k+1}\|_2}{\|x_k\|_2} = \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^{\frac{1}{2}} =: c$$

and the convergence constant, c, is close to 1.

Solutions to exercises for Part 3.

1(a). The unconstrained minimizer $-(1,0,1/2)^T$ has ℓ_2-norm $1 < \sqrt{5}/2 < 2$. Thus, since B is positive definite, the unconstrained minimizer solves the problem.

1(b). The unconstrained minimizer has too large a ℓ_2-norm, so the solution must lie on the boundary of the constraint. The solution must be of the form $-(1/(1+\lambda),0,1/(2+\lambda))^T$. To satisfy the trust-region constraint, we then must have

$$\frac{1}{(1+\lambda)^2}+\frac{1}{(2+\lambda)^2}=\Delta^2=\frac{25}{144}$$

which has a root $\lambda = 2$. Thus the required solution is $-(1/3,0,1/4)^T$.

1(c). The Hessian is indefinite so the solution must lie on the boundary of the constraint. The solution is then of the form $-(1/(-2+\lambda),0,1/(-1+\lambda))^T$. To satisfy the trust-region constraint, we then must have

$$\frac{1}{(-2+\lambda)^2}+\frac{1}{(-1+\lambda)^2}=\Delta^2=\frac{25}{144}$$

which has a root $\lambda = 5$ (c.f. the previous equation with a change of variables $\hat{\lambda} = \lambda+3$) at which $B+\lambda I$ is positive semi-definite. Thus again the solution is $-(1/3,0,1/4)^T$.

1(d). Again B is indefinite, and so the solution must be of the form $-(\omega,0,1/(-1+\lambda))^T$, where $\omega = 0/(-2+\lambda)$ can only be nonzero if $\lambda = 2$—note that $B+\lambda I$ is only positive semi-definite when $\lambda \geq 2$. Suppose that $\lambda > 2$. To satisfy the trust-region constraint, we then must have

$$\frac{1}{(-1+\lambda)^2}=\Delta^2=\frac{1}{4}$$

which has roots 1 ± 2. The desired root is $\lambda = 3$, from which we deduce the solution is $-(0,0,1/2)^T$.

1(e). As in (d), if we guess that $\lambda > 2$, we find that the roots of

$$\frac{1}{(-1+\lambda)^2}=\Delta^2=2$$

are $1\pm 1/\sqrt{2} < 2$. So λ must be 2, and the solution is of the form $-(\omega,0,1)^T$. To satisfy the trust-region constraint, we then must have

$$\omega^2+1=\Delta^2=2,$$

and hence $\omega = \pm 1$. Thus the required solution is $-(\pm 1,0,1)^T$.

Solutions to exercises for Part 4.

1(a). The first-order optimality conditions (Theorem 1.8) are that $x_2 \geq 0$ (primal feasibility),

$$\begin{pmatrix} x_1 \\ 1 \end{pmatrix} - y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$$

and $y \geq 0$ (dual feasibility), and $y \cdot x_2 = 0$ (complementary slackness). Dual feasibility says that $y = 1$ and $x_1 = 0$, from which we deduce that $x_2 = 0$ from complementary slackness. Second-order optimality conditions are simply that

$$s_1^2 = (s_1, s_2)^T \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \geq 0$$

for all $s \neq 0$ for which $s_2 = 0$ which are automatically satisfied. Thus the solution is $x = (0, 0)$ with Lagrange multiplier $y = 1$.

1(b). The logarithmic barrier function is

$$\Phi(x, \mu) = \tfrac{1}{2} x_1^2 + x_2 - \mu \log x_2.$$

The first-order optimality conditions for the unconstrained minimization of Φ are that

$$\begin{pmatrix} x_1 \\ 1 \end{pmatrix} - \mu \begin{pmatrix} 0 \\ x_2^{-1} \end{pmatrix} = 0.$$

If we let $x(\mu)$ be the desired minimizer, the optimality conditions indicate that $x(\mu) = (0, \mu)$, while the Lagrange multiplier estimates are $y(\mu) = c(x(\mu))/\mu = 1$. The Hessian is positive definite

1(c). The Hessian matrix is

$$\begin{pmatrix} 1 & 0 \\ 0 & \mu x_2^{-2} \end{pmatrix};$$

at the minimizer of $\Phi(x, \mu)$, the Hessian is

$$\begin{pmatrix} 1 & 0 \\ 0 & \mu^{-1} \end{pmatrix}.$$

The eigenvalues are 1 and μ^{-1}. As μ goes to zero, one eigenvalue diverges to infinity, while the other one stays fixed at 1.

1(d). The primal-dual system at $x(\mu)$ is

$$\begin{pmatrix} 1 & 0 \\ 0 & \mu^{-1} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = - \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} - \bar{\mu} \begin{pmatrix} 0 \\ \mu^{-1} \end{pmatrix} \right]$$

Thus $s_1 = 0$, while $s_2 = -\mu + \bar{\mu}$. In particular $x(\mu) + s = \bar{\mu} = x(\bar{\mu})$, the minimizer of $\Phi(x, \bar{\mu})$!

2(a). The logarithmic barrier function is

$$\Phi(x,\mu) = x^T g + \tfrac{1}{2} x^T B x - \mu \log(\Delta^2 - x^T x).$$

Its gradient is

$$\nabla_x \Phi(x,\mu) = g + Bx + \frac{2\mu}{\Delta^2 - x^T x} x,$$

and its Hessian is

$$\nabla_{xx} \Phi(x,\mu) = B + \frac{2\mu}{\Delta^2 - x^T x} I + \frac{2\mu}{(\Delta^2 - x^T x)^2} x x^T.$$

2(b). The first-order optimality condition is that

$$(B + \frac{2\mu}{\Delta^2 - x^T x} I) x = -g. \tag{1}$$

If we define

$$\lambda(\mu) = \frac{2\mu}{\Delta^2 - x^T x},$$

(1) is precisely the requirement

$$(B + \lambda(\mu) I) x = -g$$

from Theorem 3.9. Moreover, $\lambda(\mu) > 0$. However,

$$\lambda(\mu)(\Delta^2 - x^T x) = 2\mu$$

and we need μ to converge to zero to satisfy all of the first-order requirements in Theorem 3.9.

Solutions to exercises for Part 5.

1(a). We first need to check that $s^T Bs \geq 0$ when $As = 0$, as otherwise the solution lies at infinity. In all cases B is diagonal, so we write $B = \text{diag}(b_{11} \;\; b_{22} \;\; b_{33})$. It is easy to see that the columns of the matrix

$$N = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}$$

form a basis for the null-space of A, so we need to check that

$$N^T BN = \begin{pmatrix} b_1 + b_2 & 0 \\ 0 & b_3 \end{pmatrix}$$

is positive semi-definite. For our first example $N^T BN$ has all its eigenvalues at 1, so the minimizer is finite. The minimizer satisfies

$$\begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

which gives $x = (-2, 4, 1)$ and $y = 5$.

1(b). In this case $N^T BN$ has eigenvalues 0 and 1, so there is a solution if and only if

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

is consistent. The system gives $x_3 = 1$, but then the remaining equations lead to both $-x_2 + y = 1$ and $-x_2 + y = -1$. Thus the problem is unbounded from below.

1(c). In this case $N^T BN$ has eigenvalues -1 and 1, so the problem is unbounded from below, and the solution lies at infinity.

2. The gradient of the augmented Lagrangian function at x_k, y_k, μ_k is

$$\nabla_x \Phi(x_k) = g_k + A_k^T \left(\frac{c_k}{\mu_k} - y_k \right).$$

The SQP search direction s_k and its associated Lagrange multiplier estimates y_{k+1} satisfy

$$B_k s_k - A_k^T y_{k+1} = -g_k \tag{2}$$

and

$$A_k s_k = -c_k. \tag{3}$$

Premultiplying (2) by s_k and using (3) gives that

$$s_k^T g_k = -s_k^T B_k s_k + s_k^T A_k^T y_{k+1} = -s_k^T B_k s_k - c_k^T y_{k+1} \tag{4}$$

Likewise (3) gives

$$\frac{1}{\mu_k} s_k^T A_k^T c_k = -\frac{\|c_k\|_2^2}{\mu_k}. \tag{5}$$

Combining (4) and (5), and using the positive definiteness of B_k, the Cauchy-Schwarz inequality and the fact that $s_k \neq 0$ if x_k is not critical, yields

$$\begin{aligned} s_k^T \nabla_x \Phi(x_k) = &\ s_k^T \left[g_k + A_k^T \left(\frac{c_k}{\mu_k} - y_k \right) \right] \\ = &\ -s_k^T B_k s_k - c_k^T (y_{k+1} - y_k) - \frac{\|c_k\|_2^2}{\mu_k} \\ < &\ -\|c_k\|_2 \left(\frac{\|c_k\|_2}{\mu_k} - \|y_{k+1} - y_k\|_2 \right) \\ \leq &\ 0 \end{aligned}$$

because of the required bound on μ_k.