1(a). Given an estimate x_k of a local minimizer of $f(x)$, a *linesearch* method (i) computes a search direction s_k, which must also be a descent direction (i.e., $s_k^T \nabla_x f(x_k) < 0$), and (ii) computes a stepsize α_k so that $f(x_k + \alpha_k s_k)$ is "sufficiently" smaller than $f(x_k)$ (using for instance a backtracking Armijo rule). The next iterate is $x_{k+1} = x_k + \alpha_k s_k$.

The Armijo condition is that the stepsize α_k must satisfy

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \beta \alpha_k p_k^T \nabla_x f(x_k)$$

for some $\beta \in (0,1)$.

Let $\mathcal{N} = \{0, 1, 2, \ldots\}$. Given an initial "guess" at the stepsize α_{init} and a sequence of decreasing stepsizes $\{\alpha_{init}\tau^i\}_{i \in \mathcal{N}}$ for some $\tau \in (0,1)$, a backtracking-Armijo linesearch sets $\alpha_k = \alpha_{init}\tau^l$ where l is the smallest member $\mathcal{N}$ for which

$$f(x_k + \alpha^{(l)} p_k) \leq f(x_k) + \beta \alpha^{(l)} p_k^T \nabla_x f(x_k).$$

1(b). The Newton direction is a solution p_k to the system

$$\nabla_{xx} f(x_k) p_k = -\nabla_x f(x_k)$$

if a solution exists. This may be a poor direction because

(i) it may not give a descent (i.e., $p_k^T \nabla_x f(x_k) \geq 0$), which may happen if $\nabla_{xx} f(x_k)$ is indefinite, or

(ii) it may not give sufficient descent (i.e, $p_k^T \nabla_x f(x_k) \longrightarrow 0$ but $\nabla_x f(x_k) \not\longrightarrow 0$), which may happen if $\nabla_{xx} f(x_k)$ is not sufficiently positive definite, or

(iii) it may become too small relative to $\nabla_x f(x_k)$, which may happen if $\nabla_{xx} f(x_k) \longrightarrow \infty$

1(c). To modify the Newton direction, replace the Newton system by

$$(\nabla_{xx} f(x_k) + M_k) p_k = -\nabla_x f(x_k)$$

where M_k is chosen so that $\nabla_{xx} f(x_k) + M_k$ is "sufficiently" positive definite and $M_k = 0$ when $\nabla_{xx} f(x_k)$ is itself "sufficiently" positive definite.

There are various ways of doing this. For example, if $\nabla_{xx} f(x_k)$ has the spectral decomposition $\nabla_{xx} f(x_k) = Q_k D_k Q_k^T$, where Q_k is an orthonormal matrix of eigenvectors, and D_k a diagonal matrix of eigenvalues, then pick

$$\nabla_{xx} f(x_k) + M_k = Q_k \max(\epsilon, |D_k|) Q_k^T$$

for some small $\epsilon > 0$. Alternatively one could pick $M_k = \max(0, -\lambda_{\min}(\nabla_{xx} f(x_k)))I$ where $\lambda_{\min}(\nabla_{xx} f(x_k)$ is the smallest eigenvalue of $\nabla_{xx} f(x_k)$, or use a modified Cholesky factorization.

1(d). The gradient and Hessian of $f(x)$ are

$$\nabla_x f(x) = \begin{pmatrix} 2x_1 \\ 200x_2 \end{pmatrix} \quad \text{and} \quad \nabla_{xx} f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}$$

and thus the Newton direction at $(1, 1)$ is

$$p = -\begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 200 \end{pmatrix} = -\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This direction is a descent direction, since $\nabla_{xx} f(x)$ is positive definite (or since $p^T \nabla_x f(x) = -202 < 0$), so is suitable for a linesearch method.

Since $f(x + p) = 0$ and $f(x) + \beta p^T \nabla_x f(x) = 101 - 202\beta$ the Armijo condition will accept the unit stepsize ($\alpha = 1$) so long as

$$0 = f(x + p) \leq f(x) + \beta p^T \nabla_x f(x) = 101 - 202\beta$$

that is, provided that

$$\beta \leq \tfrac{1}{2}.$$

2(a). Second-order sufficient conditions are that there exist Lagrange multipliers y_* for which x_* is primal feasible, i.e.,

$$c(x_*) = 0,$$

and dual feasible i.e.,

$$\nabla_x f(x_*) - (\nabla c(x_*))^T y_* = 0,$$

and in addition

$$v^T \left(\nabla_{xx} f(x_*) - \sum_i (y_*)_i \nabla_{xx} c_i(x_*) \right) v > 0$$

for all nonzero v such that $(\nabla_x c(x_*))v = 0$.

2(b). For the given example, $\nabla_x c_i(x_*) = A$ and $\nabla_{xx} f(x_*) - \sum_i (y_*)_i \nabla_{xx} c_i(x_*) = B$. We first need to check that $v^T B v \geq 0$ when $Av = 0$, as otherwise the solution lies at infinity. For our example B is diagonal, so we write $B = \mathrm{diag}(b_1 \;\; b_2 \;\; b_3)$. It is easy to see [the hint] that the columns of the matrix

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

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

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

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

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

which gives $s = (1, 4, -2)$ and $w = 5$.

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

2(d). For the given example $\nabla_x g(x_*) = (2x_1 + 1, 1, 3x_3 - 2) = (1, 1, 1)$, $\nabla_x c_i(x_*) = (x_1, x_2, x_3) = (0, 1, 1)$ and

$$\nabla_{xx} f(x_*) - \sum_i (y_*)_i \nabla_{xx} c_i(x_*) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix} - 1. \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

But then this is precisely the same as example in 2(b), so the solution is $s = (1, 4, -2)$.

2(e). The trust-region constraint requires that $s_2 \leq 1/2$ and $s_3 \leq 1/2$ so that $s_2 + s_3 \leq 1$. But this is inconsistent with the linearized constraint $s_2 + s_3 = 2$, so the subproblem has no solution.