Solutions to the ACM 113 Final Exam.

1. The Baywatch Problem.
It is obvious that the paths across the sand and through the water should be straight lines, since we assume constant velocities. Thus the total path is piece-wise straight, and the question is reduced to finding the directions of the segments. Denote the angles with the vertical axes θ_1 and θ_2 for the sand and water parts of the trajectory, respectively. Also denote the horizontal projections of the pieces x_1 and x_2 and the vertical projections, r_1 and r_2 (see picture).

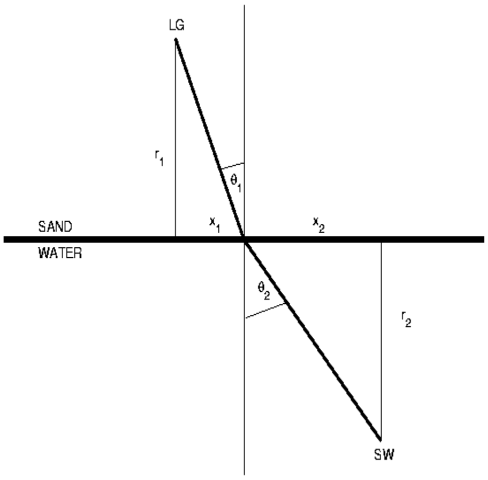

Figure 1: The Baywatch Problem

Then we can formulate our problem as the following optimization problem:

$$\begin{aligned} \min\ & \frac{r_1}{v_1 \cos\theta_1} + \frac{r_2}{v_2\cos\theta_2} \\ \text{s.t. } & x_1 + x_2 = a = \text{const} \\ & 0 \le \theta_1, \theta_2 \le \frac{\pi}{2} \end{aligned}$$

From geometry, $x_i = r_i \tan\theta_i$, $i = 1, 2$. Also, a well-known trigonometric identity implies $1/\cos\theta_i = \sqrt{1+\tan^2\theta_i}$, $i = 1, 2$. Thus, defining $t_i = \tan\theta_i$, $i = 1, 2$, we can write our optimization problem in the new variables as follows

$$\begin{aligned} \min\ & \frac{r_1}{v_1}\sqrt{1+t_1^2} + \frac{r_2}{v_2}\sqrt{1+t_2^2} \\ \text{s.t. } & r_1 t_1 + r_2 t_2 - a = 0 \\ & t_1, t_2 \ge 0 \end{aligned}$$

The Lagrangian is

$$\mathcal{L}(t,\lambda) = \frac{r_1}{v_1}\sqrt{1+t_1^2} + \frac{r_2}{v_2}\sqrt{1+t_2^2} - \lambda(r_1 t_1 + r_2 t_2 - a)$$

The first of the KKT conditions are

$$\frac{\partial \mathcal{L}}{\partial t_i} = \frac{r_i}{v_i}\frac{t_i}{\sqrt{1+t_i^2}} - \lambda r_i = 0$$

implying

$$\frac{t_i}{\sqrt{1+t_i^2}} = \lambda v_i, \quad i = 1,2$$

Recalling that $t_i = \tan\theta_i$ and invoking the trigonometric identity one more time, we observe that this expression is equivalent to $\sin\theta_i = \lambda^* v_i$, which implies

$$\frac{\sin\theta_1}{v_1} = \frac{\sin\theta_2}{v_2}$$

Thus the optimal path is a union of two straight line segments. The sines of anlges that the segments make with the vertical are proportional to the corresponding velocities. Note that this result is known as Snell's law in geometrical optics.

2. Local Convergence of Trust Region Methods

To prove part a), we consider several cases. First suppose $\nabla f_k^T B_k \nabla f_k = \nabla f_k^T \nabla^2 f_k \nabla f_k \le 0$ (note that even though $\nabla^2 f(x^*)$ is positive definite, we cannot in general assume that this property holds at every iterate x_k as well). in this case, by definition,

$$m_k(0) - m_k(p_k^c) = f_k - m_k(-\Delta_k \nabla f_k / \|\nabla f_k\|) = \frac{\Delta_k}{\|\nabla f_k\|}\|\nabla f_k\|^2 - \frac{1}{2}\frac{\Delta_k^2}{\|\nabla f_k\|^2}\nabla f_k^T B_k \nabla f_k \ge$$

$$\ge \Delta_k \|\nabla f_k\| \ge \|\nabla f_k\| \min\left(\Delta_k, \frac{\|\nabla f_k\|}{\|B_k\|}\right) \ge \frac{1}{2}\|\nabla f_k\| \min\left(\Delta_k, \frac{\|\nabla f_k\|}{\|\nabla^2 f_k\|}\right)$$

as required.

Now consider the case when $\nabla f_k^T B_k \nabla f_k = \nabla f_k^T \nabla^2 f_k \nabla f_k > 0$. Recall that two expressions for the Cauchy point are possible in this case, depending on whether

$$\frac{\|\nabla f_k\|^3}{\Delta_k \nabla f_k^T B_k \nabla f_k} \le 1 \tag{1}$$

holds or not. Suppose (1) holds. Then we have $\tau_k = \|\nabla f_k\|^3 / \Delta_k \nabla f_k^T B_k \nabla f_k$ and so

$$m_k(0) - m_k(p_k^c) = \frac{\|\nabla f_k\|^4}{\nabla f_k^T B_k \nabla f_k} - \frac{1}{2}\nabla f_k^T B_k \nabla f_k \frac{\|\nabla f_k\|^4}{(\nabla f_k^T B_k \nabla f_k)^2} =$$

$$= \frac{1}{2}\frac{\|\nabla f_k\|^4}{\nabla f_k^T B_k \nabla f_k} \ge \frac{\|\nabla f_k\|^4}{\|B_k\|\,\|\nabla f_k\|^2} = \frac{1}{2}\frac{\|\nabla f_k\|^2}{\|B_k\|} \ge$$

$$\ge \frac{1}{2}\|\nabla f_k\| \min\left(\Delta_k, \frac{\|\nabla f_k\|^2}{\|B_k\|}\right) = \frac{1}{2}\|\nabla f_k\| \min\left(\Delta_k, \frac{\|\nabla f_k\|^2}{\|\nabla^2 f_k\|}\right)$$

Finally, if (1) does not hold, i.e.,

$$\nabla f_k^T B_k \nabla f_k = \nabla f_k^T \nabla^2 f_k \nabla f_k < \frac{\|\nabla f_k\|^3}{\Delta_k}$$

we have $\tau_k = 1$ by definition. Using these two identities, we deduce that

$$m_k(0) - m_k(p_k^c) = \frac{\Delta_k}{\|\nabla f_k\|}\|\nabla f_k\|^2 - \frac{1}{2}\frac{\Delta_k^2}{\|\nabla f_k\|^2}\nabla f_k^T B_k \nabla f_k \ge$$

$$\ge \Delta_k \|\nabla f_k\| - \frac{1}{2}\frac{\Delta_k^2}{\|\nabla f_k\|^2}\frac{\|\nabla f_k\|^3}{\Delta_k} = \frac{1}{2}\Delta_k\|\nabla f_k\| \ge$$

$$\ge \frac{1}{2}\|\nabla f_k\| \min\left(\Delta_k, \frac{\|\nabla f_k\|^2}{\|B_k\|}\right) = \frac{1}{2}\|\nabla f_k\| \min\left(\Delta_k, \frac{\|\nabla f_k\|^2}{\|\nabla^2 f_k\|}\right)$$

Thus part a) holds for all possible cases.

To prove part b), observe that, since $p_k^N = -\nabla^2 f_k^{-1} \nabla f_k$, we have

$$\|p_k - p_k^N\| \le \|\nabla^2 f(x_k)^{-1}\| \, \|\nabla^2 f(x_k) p_k - \nabla^2 f(x_k) p_k^N\| = \|\nabla^2 f(x_k)^{-1}\| \, \|r_k\| \le$$
$$\le \eta_k \|\nabla^2 f(x_k)^{-1}\| \, \|\nabla f_k\| \le \eta_k \|\nabla^2 f(x_k)^{-1}\| \, \|\nabla^2 f(x_k)\| \, \|p_k^N\| \le \kappa_{\nabla^2 f(x_k)} \eta_k \|p_k^N\|$$

Since the condition number of $\nabla^2 f(x_k)$ is for k large enough, the right-hand side is $o(\|p_k^N\|)$ if only $\eta_k \to 0$, so b) is proved.

3. LP Sensitivity analysis.

a) If $\tilde{b} = b + \Delta b$, then $\tilde{x}_B = B^{-1}\tilde{b}$. Since feasibility requires $\tilde{x}_B \ge 0$, the current basis is still feasible if $x_B \ge -B^{-1}\Delta b$. In this case the new values of the basic variables are $\bar{x}_B = x_B + B^{-1}\Delta b$ and the new optimal value of the objective is $\bar{z} = z + y^T \Delta b$, where $z = c_B^T x_B$ and y is the vector of dual variables. There are no changes to either y or the dual slack variables s. If $x_B < -B^{-1}\Delta b$, the new basis is infeasible, and the dual simplex method can be used to restore feasibility.

b) If $\tilde{c}_N = c_N + \Delta c_N$, then current basis will still be optimal if $\tilde{s}_N = s_N + \Delta c_N \ge 0$, i.e., when $s_N \ge -\Delta c_N$. In this case, there are no changes to x, y or z. If $s_N < -\Delta c_N$, then dual feasibility is violated, and the primal simplex method can be used to restore optimality.

c) If $\tilde{N} = N + \Delta N$, then the current basis will remain optimal if $\tilde{s}_N^T = s_N^T - c_B^T B^{-1} \Delta N \ge 0$, i.e., $s_N^T \ge c_B^T B^{-1} \Delta N$. In this case x, y and z are unchanged. If this condition is not satisfied, then dual feasibility is violated, and the optimality can be restored by applying the primal simplex method.

d) If x_t is a new variable with the specified objective and constraint coefficients, the current basis stays optimal if the corresponding new dual slack variable s_t is nonnegative, i.e., if $s_t = c_t - y^T A_t \ge 0$. If this holds, then there are no changes to the objective or any of the variables, except that $s_t = c_t - y^T A_t$ will be added to s. If $c_t - y^T A_t < 0$, then the solution is dual infeasible, and applying the primal simplex method will restore optimality.

4. Augmented Lagrangian Methods

a) The augmented Lagrangian and its gradient are given by

$$\mathcal{L}_a(x, \lambda) = e^{x_1 x_2 x_3 x_4 x_5} - \lambda_1 (x^T x - 10) - \lambda_2 (x_2 x_3 - 5 x_4 x_5) - \lambda_3 (x_1^3 + x_2^3 + 1) +$$
$$+ \frac{a}{2}((x^T x - 10)^2 + (x_2 x_3 - 5 x_4 x_5)^2 + (x_1^3 + x_2^3 + 1)^2)$$

$$\nabla_x \mathcal{L}(x, \lambda) = (\mathcal{L}_{a1}, \mathcal{L}_{a2}, \mathcal{L}_{a3}, \mathcal{L}_{a4}, \mathcal{L}_{a5})^T, \quad \text{where}$$

$$\mathcal{L}_{a1} \equiv \frac{\partial \mathcal{L}_a}{\partial x_1} = x_2 x_3 x_4 x_5 e^{x_1 x_2 x_3 x_4 x_5} + x_1(-2\lambda_1 + a(x^T x - 10)) +$$
$$+ 3x_1^2(-\lambda_3 + a(x_1^3 + x_2^3 + 1))$$
$$\mathcal{L}_{a2} \equiv \frac{\partial \mathcal{L}_a}{\partial x_2} = x_1 x_3 x_4 x_5 e^{x_1 x_2 x_3 x_4 x_5} + x_2(-2\lambda_1 + a(x^T x - 10)) +$$

$$+x_3(-\lambda_2 + a(x_2x_3 - 5x_4x_5)) + 3x_2^2(-\lambda_3 + a(x_1^3 + x_2^3 + 1))$$

$$\mathcal{L}_{a3} \equiv \frac{\partial \mathcal{L}_a}{\partial x_3} = x_1x_2x_4x_5e^{x_1x_2x_3x_4x_5} + x_3(-2\lambda_1 + a(x^Tx - 10)) + \\ +x_2(-\lambda_2 + a(x_2x_3 - 5x_4x_5))$$

$$\mathcal{L}_{a4} \equiv \frac{\partial \mathcal{L}_a}{\partial x_4} = x_1x_2x_3x_5e^{x_1x_2x_3x_4x_5} + x_4(-2\lambda_1 + a(x^Tx - 10)) + \\ +5x_5(-\lambda_2 + a(x_2x_3 - 5x_4x_5))$$

$$\mathcal{L}_{a5} \equiv \frac{\partial \mathcal{L}_a}{\partial x_5} = x_1x_2x_3x_4e^{x_1x_2x_3x_4x_5} + x_5(-2\lambda_1 + a(x^Tx - 10)) + \\ +5x_4(-\lambda_2 + a(x_2x_3 - 5x_4x_5))$$

b) Attached source code presents the implementation of the augmented Lagrangian method with Hessian-free finite-difference truncated Newton unconstrained sub-iterations. The parameters of the sub-iterates were chosen to ensure superlinear convergence ($\eta_k = \min\{1/2, \sqrt{\|r_k\|}\}$). Initial conditions were taken from part c). The optimal solution is $f^* = 0.053950$ at the point $x^* = (-1.71714, 1.59571, 1.82725, -0.76364, -0.76364)^T$ with $\lambda^* = (-0.04016, 0.03796, -0.00522)^T$.

c) When $a > \bar{a}$ is held constant, the convergence of the outer iterations is linear, but the unconstrained sub-problems are easier to solve. When $a \to \infty$, the onvergence of the outer iterations approaches superlinear, but the unconstrained subproblems are harder to solve as they become ill-conditioned. This is the behavior that the given problem showed: the first case required more outer iterations, but the sub-problems actually converged in fewer iterations. In the tables below we demonstrate convergence of the methods for the two cases.

iterate	x_1	x_2	x_3	x_4	x_5	total iters.
$k=0$	-2	2	2	-1	-1	-
$k=1$	-1.717352795	1.596019915	1.827374513	-0.764238266	-0.764238266	14
$k=2$	-1.717143656	1.595709829	1.827245784	-0.763643396	-0.763643396	7
$k=3$	-1.717143591	1.595709714	1.827245715	-0.763643076	-0.763643076	3
result	-1.717143591	1.595709714	1.827245715	-0.763643076	-0.763643076	24

iterate	λ_1	λ_2	λ_3	total iters.
$k=0$	0	0	0	-
$k=1$	-0.039980620	0.037745202	-0.005192523	14
$k=2$	-0.040162652	0.037957664	-0.005222628	7
$k=3$	-0.040162745	0.037957775	-0.005222643	3
result	-0.040162745	0.037957775	-0.005222643	24

Table 1: Convergence for $a \to \infty$.

iterate	x_1	x_2	x_3	x_4	x_5	total iters
$k=0$	-2	2	2	-1	-1	-
$k=1$	-1.717352795	1.596019915	1.827374514	-0.764238266	-0.764238266	14
$k=2$	-1.717144417	1.595711063	1.827246072	-0.763646243	-0.763646243	5
$k=3$	-1.717143575	1.595709697	1.827245754	-0.763643094	-0.763643094	2
$k=4$	-1.717143571	1.595709691	1.827245751	-0.763643078	-0.763643078	1
result	-1.717143571	1.595709691	1.827245751	-0.763643078	-0.763643078	22

iterate	λ_1	λ_2	λ_3	total iters.
$k=0$	0	0	0	-
$k=1$	-0.039980620	0.037745202	-0.005192523	14
$k=2$	-0.040161824	0.037956677	-0.005222489	5
$k=3$	-0.040162740	0.037957769	-0.005222643	2
$k=4$	-0.040162745	0.037957774	-0.005222643	1
result	-0.040162745	0.037957774	-0.005222643	22

Table 2: Convergence for $a = \bar{a}$ =const.

5. Protein Design is NP-complete

a) PRODES in decision form: Does there exist a side chain at each position such that $E_{\text{total}} \leq C$?
b) PRODES is in NP since a "yes" instance can be represented concisely by p side chain indices and verified in polynomial-time by evaluating E_{total}.
c) Consider SATISFIABILITY with M clauses and N variables, where clause i has L_i literals. Construct PRODES with $p = M + N$ positions:

- M positions with $n_i = L_i$ side-chains each representing a literal (either a variable or its negation),
- N positions with $n_i = 2$ side-chains each representing T or F.

Assign energies to side-chain pairs:

- $E(i_r, j_s) = E(j_s, i_r) = 2C$ if i_r represents a variable and j_s represents F,
- $E(i_r, j_s) = E(j_s, i_r) = 2C$ if i_r represents a variable negation and j_s represents T,
- $E(i_r, j_s) = 0$ otherwise.

With this specification, $E_{\text{total}} = 0$ if and only if the original problem is satisfiable. Hence, an instance of PRODES can be constructed in polynomial-time given an instance of SATISFIABILITY and the instance will be "yes" for PRODES if and only if it is "yes" for SATISFIABILITY. Hence, PRODES is NP-complete.