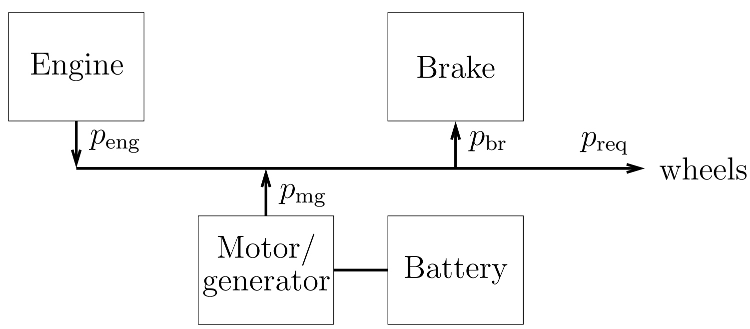
EE364a Homework 5 solutions

4.65 *Optimal operation of a hybrid vehicle.* A hybrid vehicle has an internal combustion engine, a motor/generator connected to a storage battery, and a conventional (friction) brake. In this exercise we consider a (highly simplified) model of a *parallel hybrid vehicle*, in which both the motor/generator and the engine are directly connected to the drive wheels. The engine can provide power to the wheels, and the brake can take power from the wheels, turning it into heat. The motor/generator can act as a motor, when it uses energy stored in the battery to deliver power to the wheels, or as a generator, when it takes power from the wheels or engine, and uses the power to charge the battery. When the generator takes power from the wheels and charges the battery, it is called *regenerative braking*; unlike ordinary friction braking, the energy taken from the wheels is *stored*, and can be used later. The vehicle is judged by driving it over a known, fixed test track to evaluate its fuel efficiency.

A diagram illustrating the power flow in the hybrid vehicle is shown below. The arrows indicate the direction in which the power flow is considered positive. The engine power p_{eng}, for example, is positive when it is delivering power; the brake power p_{br} is positive when it is taking power from the wheels. The power p_{req} is the required power at the wheels. It is positive when the wheels require power (*e.g.*, when the vehicle accelerates, climbs a hill, or cruises on level terrain). The required wheel power is negative when the vehicle must decelerate rapidly, or descend a hill.



All of these powers are functions of time, which we discretize in one second intervals, with $t = 1, 2, \ldots, T$. The required wheel power $p_{\mathrm{req}}(1), \ldots, p_{\mathrm{req}}(T)$ is given. (The speed of the vehicle on the track is specified, so together with known road slope information, and known aerodynamic and other losses, the power required at the wheels can be calculated.)

Power is conserved, which means we have

$$p_{\mathrm{req}}(t) = p_{\mathrm{eng}}(t) + p_{\mathrm{mg}}(t) - p_{\mathrm{br}}(t), \quad t = 1, \ldots, T.$$

The brake can only dissipate power, so we have $p_{\text{br}}(t) \geq 0$ for each t. The engine can only provide power, and only up to a given limit $P_{\text{eng}}^{\max}$, *i.e.*, we have

$$0 \leq p_{\text{eng}}(t) \leq P_{\text{eng}}^{\max}, \quad t = 1, \ldots, T.$$

The motor/generator power is also limited: p_{mg} must satisfy

$$P_{\text{mg}}^{\min} \leq p_{\text{mg}}(t) \leq P_{\text{mg}}^{\max}, \quad t = 1, \ldots, T.$$

Here $P_{\text{mg}}^{\max} > 0$ is the maximum motor power, and $-P_{\text{mg}}^{\min} > 0$ is the maximum generator power.

The battery charge or energy at time t is denoted $E(t)$, $t = 1, \ldots, T+1$. The battery energy satisfies

$$E(t+1) = E(t) - p_{\text{mg}}(t) - \eta |p_{\text{mg}}(t)|, \quad t = 1, \ldots, T,$$

where $\eta > 0$ is a known parameter. (The term $-p_{\text{mg}}(t)$ represents the energy removed or added the battery by the motor/generator, ignoring any losses. The term $-\eta|p_{\text{mg}}(t)|$ represents energy lost through inefficiencies in the battery or motor/generator.)

The battery charge must be between 0 (empty) and its limit $E_{\text{batt}}^{\max}$ (full), at all times. (If $E(t) = 0$, the battery is fully discharged, and no more energy can be extracted from it; when $E(t) = E_{\text{batt}}^{\max}$, the battery is full and cannot be charged.) To make the comparison with non-hybrid vehicles fair, we fix the initial battery charge to equal the final battery charge, so the net energy change is zero over the track: $E(1) = E(T+1)$. We do not specify the value of the initial (and final) energy.

The objective in the problem is the total fuel consumed by the engine, which is

$$F_{\text{total}} = \sum_{t=1}^{T} F(p_{\text{eng}}(t)),$$

where $F : \mathbf{R} \to \mathbf{R}$ is the *fuel use characteristic* of the engine. We assume that F is positive, increasing, and convex.

Formulate this problem as a convex optimization problem, with variables $p_{\text{eng}}(t)$, $p_{\text{mg}}(t)$, and $p_{\text{br}}(t)$ for $t = 1, \ldots, T$, and $E(t)$ for $t = 1, \ldots, T+1$. Explain why your formulation is equivalent to the problem described above.

Solution. We first collect the given objective and constraints to form the problem

$$\begin{array}{ll}
\text{minimize} & \sum_{t=1}^{T} F(p_{\text{eng}}(t)) \\
\text{subject to} & p_{\text{req}}(t) = p_{\text{eng}}(t) + p_{\text{mg}}(t) - p_{\text{br}}(t) \\
& E(t+1) = E(t) - p_{\text{mg}}(t) - \eta |p_{\text{mg}}(t)| \\
& 0 \leq E(t) \leq E_{\text{batt}}^{\max} \\
& E(1) = E(T+1) \\
& 0 \leq p_{\text{eng}}(t) \leq P_{\text{eng}}^{\max} \\
& P_{\text{mg}}^{\min} \leq p_{\text{mg}}(t) \leq P_{\text{mg}}^{\max} \\
& 0 \leq p_{\text{br}}(t),
\end{array}$$

where each constraint is imposed for the appropriate range of t. The fuel use function F is convex, so the objective function is convex. With the exception of the battery charge equations, each constraint is a linear equality or linear inequality. So in this form the problem is *not* convex.

We need to show how to deal with the nonconvex constraints

$$E(t+1) = E(t) - p_{\mathrm{mg}}(t) - \eta|p_{\mathrm{mg}}(t)|.$$

One approach is to replace this constraint with the relaxation,

$$E(t+1) \leq E(t) - p_{\mathrm{mg}}(t) - \eta|p_{\mathrm{mg}}(t)|,$$

which is convex, in fact, two linear inequalities. Intuitively, this relaxation means that we open the possibility of throwing energy from the battery away at each step. This sounds like a bad idea, when fuel efficiency is the goal, and indeed, it is easy to see that if we solve the problem with the relaxed battery charge constraints, the optimal $E^\star$ satisfies

$$E^\star(t+1) = E^\star(t) - p_{\mathrm{mg}}(t) - \eta|p_{\mathrm{mg}}(t)|,$$

and therefore solves the original problem. To argue formally that this is the case, suppose that the solution of the relaxed problem does throw away some energy at some step t. We then construct a new trajectory, where we do not throw away the extra energy, and instead, use the energy to power the wheels, and reduce the engine power. This reduces the fuel consumption since the fuel consumption characteristic is increasing, which shows that the original could not have been optimal.

5.13 *Lagrangian relaxation of Boolean LP.* A *Boolean linear program* is an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0,1\}, \quad i = 1,\ldots,n, \end{array}$$

and is, in general, very difficult to solve. In exercise 4.15 we studied the LP relaxation of this problem,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & 0 \leq x_i \leq 1, \quad i = 1,\ldots,n, \end{array} \tag{1}$$

which is far easier to solve, and gives a lower bound on the optimal value of the Boolean LP. In this problem we derive another lower bound for the Boolean LP, and work out the relation between the two lower bounds.

(a) *Lagrangian relaxation.* The Boolean LP can be reformulated as the problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i(1 - x_i) = 0, \quad i = 1,\ldots,n, \end{array}$$

which has quadratic equality constraints. Find the Lagrange dual of this problem. The optimal value of the dual problem (which is convex) gives a lower bound on the optimal value of the Boolean LP. This method of finding a lower bound on the optimal value is called *Lagrangian relaxation.*

(b) Show that the lower bound obtained via Lagrangian relaxation, and via the LP relaxation (1), are the same. *Hint.* Derive the dual of the LP relaxation (1).

Solution.

(a) The Lagrangian is

$$\begin{aligned} L(x, \mu, \nu) &= c^T x + \mu^T (Ax - b) - \nu^T x + x^T \mathbf{diag}(\nu) x \\ &= x^T \mathbf{diag}(\nu) x + (c + A^T \mu - \nu)^T x - b^T \mu. \end{aligned}$$

Minimizing over x gives the dual function

$$g(\mu, \nu) = \begin{cases} -b^T \mu - (1/4) \sum_{i=1}^n (c_i + a_i^T \mu - \nu_i)^2 / \nu_i & \nu \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

where a_i is the ith column of A, and we adopt the convention that $a^2/0 = \infty$ if $a \neq 0$, and $a^2/0 = 0$ if $a = 0$.

The resulting dual problem is

$$\begin{array}{ll} \text{maximize} & -b^T \mu - (1/4) \sum_{i=1}^n (c_i + a_i^T \mu - \nu_i)^2 / \nu_i \\ \text{subject to} & \nu \succeq 0, \quad \mu \succeq 0. \end{array}$$

In order to simplify this dual, we optimize analytically over ν, by noting that

$$\begin{aligned} \sup_{\nu_i \geq 0} \left(-\frac{(c_i + a_i^T \mu - \nu_i)^2}{\nu_i} \right) &= \begin{cases} 4(c_i + a_i^T \mu) & c_i + a_i^T \mu \leq 0 \\ 0 & c_i + a_i^T \mu \geq 0 \end{cases} \\ &= \min\{0, 4(c_i + a_i^T \mu)\}. \end{aligned}$$

This allows us to eliminate ν from the dual problem, and simplify it as

$$\begin{array}{ll} \text{maximize} & -b^T \mu + \sum_{i=1}^n \min\{0, c_i + a_i^T \mu\} \\ \text{subject to} & \mu \succeq 0. \end{array}$$

(b) We follow the hint. The Lagrangian and dual function of the LP relaxation are

$$\begin{aligned} L(x, u, v, w) &= c^T x + u^T (Ax - b) - v^T x + w^T (x - \mathbf{1}) \\ &= (c + A^T u - v + w)^T x - b^T u - \mathbf{1}^T w \\ g(u, v, w) &= \begin{cases} -b^T u - \mathbf{1}^T w & A^T u - v + w + c = 0 \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The dual problem is

$$\begin{array}{ll} \text{maximize} & -b^T u - \mathbf{1}^T w \\ \text{subject to} & A^T u - v + w + c = 0 \\ & u \succeq 0, v \succeq 0, w \succeq 0, \end{array}$$

which is equivalent to the Lagrange relaxation problem derived above. We conclude that the two relaxations give the same value.

6.2 *ℓ_1-, ℓ_2-, and ℓ_∞-norm approximation by a constant vector.* What is the solution of the norm approximation problem with one scalar variable $x \in \mathbf{R}$,

$$\text{minimize} \quad \|x\mathbf{1} - b\|,$$

for the ℓ_1-, ℓ_2-, and ℓ_∞-norms?

Solution.

(a) ℓ_2-norm: the average $\mathbf{1}^T b/m$.

(b) ℓ_1-norm: the (or a) median of the coefficients of b.

(c) ℓ_∞-norm: the midrange point $(\max b_i - \min b_i)/2$.

Solutions to additional exercises

1. *Optimal operation of a hybrid vehicle.* Solve the instance of the hybrid vehicle operation problem described in exercise 4.65 in *Convex Optimization*, with problem data given in the file `hybrid_veh_data.m`, and fuel use function $F(p) = p + \gamma p^2$ (for $p \geq 0$).

   Hint. You will actually formulate and solve a *relaxation* of the original problem. You may find that some of the equality constraints you relaxed to inequality constraints do not hold for the solution found. This is not an error: it just means that there is no incentive (in terms of the objective) for the inequality to be tight. You can fix this in (at least) two ways. One is to go back and adjust certain variables, without affecting the objective and maintaining feasibility, so that the relaxed constraints hold with equality. Another simple method is to add to the objective a term of the form

   $$\epsilon \sum_{t=1}^{T} \max\{0, -P_{\text{mg}}(t)\},$$

   where ϵ is small and positive. This makes it more attractive to use the brakes to extract power from the wheels, even when the battery is (or will be) full (which removes any fuel incentive).

   Find the optimal fuel consumption, and compare to the fuel consumption with a non-hybrid version of the same vehicle (*i.e.*, one without a battery). Plot the braking power, engine power, motor/generator power, and battery energy versus time.

   How would you use optimal dual variables for this problem to find $\partial F_{\text{total}}/\partial E_{\text{batt}}^{\max}$, *i.e.*, the partial derivative of optimal fuel consumption with respect to battery capacity? (You can just assume that this partial derivative exists.) You do not have to give a long derivation or proof; you can just state how you would fine this derivative from optimal dual variables for the problem. Verify your method numerically, by changing the battery capacity a small amount and re-running the optimization, and comparing this to the prediction made using dual variables.

   Solution.

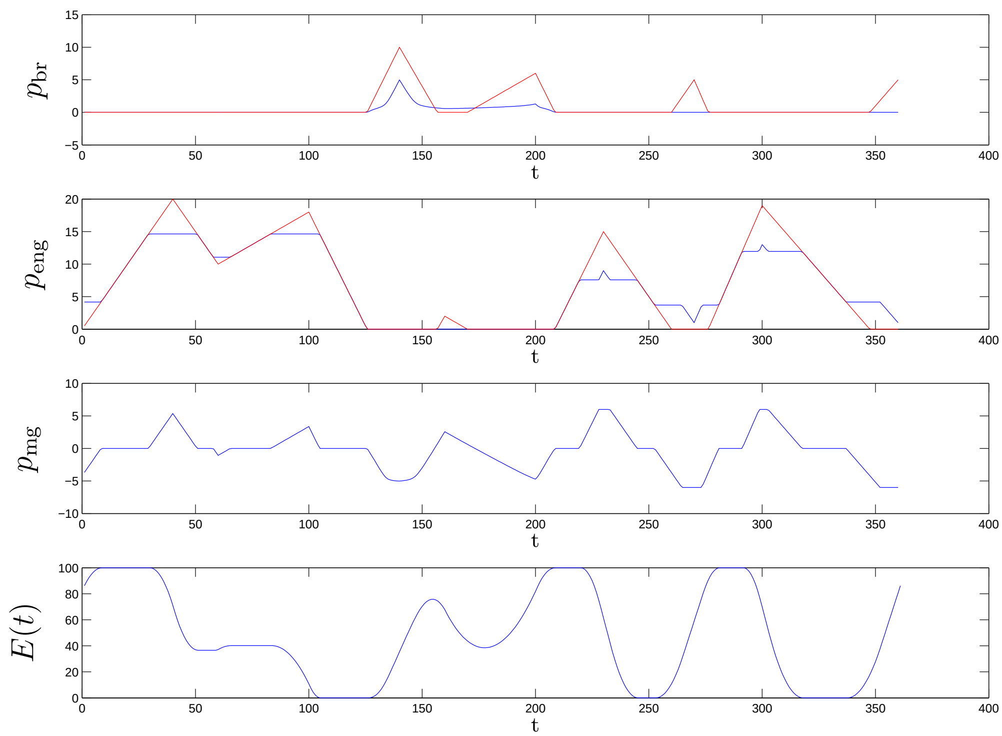
   The code for solving this instance is given below. The optimal fuel consumption of the vehicle with the battery is $F_{\text{total}} = 5077.53$. Without the battery the fuel consumption goes up to $F_{\text{total}} = 5896.81$. The battery can shift energy in time and recover energy that would have been lost in friction braking, so we expect the fuel use to be smaller than without a battery.

   Closely examining the optimal variables reveals that there is a short period when the relaxed battery energy conservation constraint *does not hold*: We actually elect to throw battery energy away. This is a brief period before $t \approx 200$, when the battery is already full, so no further charging will make any difference at all. This is not a problem: The optimal fuel consumption is correct. The constraint can be made tight by following the method in the hint, without affecting fuel optimality.

To find $\partial F_{\text{total}}/\partial E_{\text{batt}}^{\max}$, we let $\lambda^\star(t)$ be an optimal dual variable associated with the constraint $E(t) \le E_{\text{batt}}^{\max}$. We interpret $-\lambda^\star(t)$ as the partial derivative of optimal fuel use with respect to $E_{\text{batt}}^{\max}$ at time t. It tells us how the fuel usage would improve if, at time t only, we were able to store more energy in the battery. To get the full partial derivative $\partial F_{\text{total}}/\partial E_{\text{batt}}^{\max}$, we sum:

$$\frac{\partial F_{\text{total}}}{\partial E_{\text{batt}}^{\max}} = -\sum_{t=1}^{T} \lambda^\star(t).$$

This formula gives us $\partial F_{\text{total}}/\partial E_{\text{batt}}^{\max} = -4.9680$. We can verify this numerically by perturbing the constraint by a small amount and running the optimization again. Setting $E(t) \le E_{\text{batt}}^{\max} + \Delta E$, $\Delta E = 0.1$ we obtain $\partial F_{\text{total}}/\partial E_{\text{bat}}^{\max} \approx \frac{\Delta F}{\Delta E} = -4.9658$, where ΔF is the difference in fuel comsumption between the perturbed and unperturbed cases. The numerical answer is indeed very close to the answer obtained from the dual variables.




```
% instance of hybrid vehicle optimization problem,
% exercise 4.65 in Boyd & Vandenberghe, Convex Optimization
```

```
hybrid_veh_data

% solution via CVX

epsilon=1e-5;
cvx_begin
cvx_quiet(true)
variables Peng(T) Pbr(T) Pmg(T) E(T+1)
dual variable lambda
% minimize (sum(Peng+gamm*square(Peng)))  % total fuel use
minimize (sum(Peng+gamm*square(Peng))+epsilon*sum(pos(-Pmg)))
Preq == Peng + Pmg - Pbr;  % power balance
E(1) == E(T+1);  % starting and ending battery energy match
Pmg_min <= Pmg; % maximum generator power
Pmg <= Pmg_max; % maximum motor power
0 <= E;
% assign the dual variable to this set of constraints
% for use in sensitivity analysis
lambda: E <= Ebatt_max;  % battery capacity
0 <= Peng;
Peng <= Peng_max;  % max engine power
Pbr >= 0;
for t=1:T
    E(t+1)<=E(t)-Pmg(t)-eta*abs(Pmg(t));
    % this is a relaxation; true formula is below
    % E(t+1)==E(t)-Pmg(t)-eta*abs(Pmg(t));
    % extra term above is used to gaurantee tightness at optimal
end
cvx_end
Ftot = cvx_optval

% verify that the constraint holds with equality
 DE=E(2:T+1)-E(1:T)+Pmg(1:T)+eta*abs(Pmg(1:T));
 if max(abs(DE))<1e-3
     fprintf('battery energy constraint holds with equality \n')
 end

% solve the problem for a vehicle without a battery
% can do this analytically, but no harm in using CVX
cvx_begin
cvx_quiet(true)
variables Peng_nobatt(T) Pbr_nobatt(T)
```

```
minimize (sum(Peng_nobatt+gamm*square(Peng_nobatt))) % total fuel use
Preq == Peng_nobatt - Pbr_nobatt;
0 <= Peng_nobatt;
Peng_nobatt <= Peng_max;
Pbr_nobatt >= 0;
cvx_end
Ftot_no_batt = cvx_optval

% generate the plots
scrsz = get(0,'ScreenSize');
figure('Position',[1 1 scrsz(3) scrsz(4)])
subplot(4,1,1);
plot(Pbr); hold on; plot(Pbr_nobatt,'r')
title('Brake power'); xlabel('time'); ylabel('Pbr');
%legend('With battery','Without battery')
subplot(4,1,2);
plot(Peng); hold on; plot(Peng_nobatt, 'r')
title('Engine power'); xlabel('time'); ylabel('Peng');
%legend('With battery','Without battery')
subplot(4,1,3);
plot(Pmg)
title('Motor/generator power'); xlabel('time'); ylabel('Pmg');
subplot(4,1,4);
plot(E)
title('Battery stored energy'); xlabel('time'); ylabel('Ebatt');

% numerical verification of sensitivity analysis
% perturb Ebatt_max by deltaE and examine the change in F
deltaE=0.1;
cvx_begin
cvx_quiet(true)
variables Peng(T) Pbr(T) Pmg(T) E(T+1)
% minimize (sum(Peng+gamm*square(Peng)))  % total fuel use
minimize (sum(Peng+gamm*square(Peng))+epsilon*sum(pos(-Pmg)))
Preq == Peng + Pmg - Pbr;
E(1) == E(T+1);
Pmg_min <= Pmg;
Pmg <= Pmg_max;
0 <= E;
E <= Ebatt_max+deltaE;
0 <= Peng;
Peng <= Peng_max;
```

```
Pbr >= 0;
for t=1:T
    E(t+1)<=E(t)-Pmg(t)-eta*abs(Pmg(t));
end
cvx_end
Ftot_deltaE=cvx_optval;
% calculate sensitivity numerically
dFdE_num=(Ftot_deltaE-Ftot)/deltaE
% take the negative of the sum of the dual variables
% to obtain dFtot/dEbatt_max
dFdE_tot=-sum(lambda)
```

2. *Heuristic suboptimal solution for Boolean LP.* This exercise builds on exercises 4.15 and 5.13, which involve the Boolean LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0,1\}, \quad i = 1, \ldots, n, \end{array}$$

with optimal value $p^\star$. Let x^{rlx} be a solution of the LP relaxation

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & 0 \preceq x \preceq \mathbf{1}, \end{array}$$

so $L = c^T x^{\text{rlx}}$ is a lower bound on $p^\star$. The relaxed solution x^{rlx} can also be used to guess a Boolean point $\hat{x}$, by rounding its entries, based on a threshold $t \in [0,1]$:

$$\hat{x}_i = \begin{cases} 1 & x_i^{\text{rlx}} \geq t \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, \ldots, n$. Evidently $\hat{x}$ is Boolean (*i.e.*, has entries in $\{0,1\}$). If it is feasible for the Boolean LP, *i.e.*, if $A\hat{x} \preceq b$, then it can be considered a guess at a good, if not optimal, point for the Boolean LP. Its objective value, $U = c^T \hat{x}$, is an upper bound on $p^\star$. If U and L are close, then $\hat{x}$ is nearly optimal; specifically, $\hat{x}$ cannot be more than $(U - L)$-suboptimal for the Boolean LP.

This rounding need not work; indeed, it can happen that for all threshold values, $\hat{x}$ is infeasible. But for some problem instances, it can work well.

Of course, there are many variations on this simple scheme for (possibly) constructing a feasible, good point from x^{rlx}.

Finally, we get to the problem. Generate problem data using

```
rand('state',0);
n=100;
m=300;
A=rand(m,n);
b=A*ones(n,1)/2;
c=-rand(n,1);
```

You can think of x_i as a job we either accept or decline, and $-c_i$ as the (positive) revenue we generate if we accept job i. We can think of $Ax \preceq b$ as a set of limits on m resources. A_{ij}, which is positive, is the amount of resource i consumed if we accept job j; b_i, which is positive, is the amount of resource i available.

Find a solution of the relaxed LP and examine its entries. Note the associated lower bound L. Carry out threshold rounding for (say) 100 values of t, uniformly spaced over $[0, 1]$. For each value of t, note the objective value $c^T\hat{x}$ and the maximum constraint violation $\max_i(A\hat{x} - b)_i$. Plot the objective value and the maximum violation versus t. Be sure to indicate on the plot the values of t for which $\hat{x}$ is feasible, and those for which it is not.

Find a value of t for which $\hat{x}$ is feasible, and gives minimum objective value, and note the associated upper bound U. Give the gap $U - L$ between the upper bound on $p^\star$ and the lower bound on $p^\star$. If you define vectors `obj` and `maxviol`, you can find the upper bound as `U=min(obj(find(maxviol<=0)))`.

Solution.

The following Matlab code finds the solution

```
% generate data for boolean LP relaxation & heuristic
rand('state',0);
n=100;
m=300;
A=rand(m,n);
b=A*ones(n,1)/2;
c=-rand(n,1);

% solve LP relaxation
cvx_begin
   variable x(n)
   minimize (c'*x)
   subject to
       A*x <= b
       x>=0
       x<=1
cvx_end
```

```
xrlx = x;
L=cvx_optval;

% sweep over threshold & round
thres=0:0.01:1;
maxviol = zeros(length(thres),1);
obj = zeros(length(thres),1);
for i=1:length(thres)
   xhat = (xrlx>=thres(i));
   maxviol(i) = max(A*xhat-b);
   obj(i) = c'*xhat;
end

% find least upper bound and associated threshold
i_feas=find(maxviol<=0);
U=min(obj(i_feas))
%U=min(obj(find(maxviol <=0)))
t=min(i_feas);
min_thresh=thres(t)

% plot objective and max violation versus threshold
subplot(2,1,1)
plot(thres(1:t-1),maxviol(1:t-1),'r',thres(t:end),maxviol(t:end),'b','linewidth',2)
xlabel('threshold');
ylabel('max violation');
subplot(2,1,2)
hold on; plot(thres,L*ones(size(thres)),'k','linewidth',2);
plot(thres(1:t-1),obj(1:t-1),'r',thres(t:end),obj(t:end),'b','linewidth',2);
xlabel('threshold');
ylabel('objective');
```

The lower bound found from the relaxed LP is $L = -33.1672$. We find that the threshold value $t = 0.6006$ gives the best (smallest) objective value for feasible $\hat{x}$: $U = -32.4450$. The difference is 0.7222. So $\hat{x}$, with $t = 0.6006$, can be no more than 0.7222 suboptimal, *i.e.*, around 2.2% suboptimal.

In figure 1, the red lines indicate values for thresholding values which give infeasible $\hat{x}$, and the blue lines correspond to feasible $\hat{x}$. We see that the maximum violation decreases as the threshold is increased. This occurs because the constraint matrix A only has nonnegative entries. At a threshold of 0, all jobs are selected, which is an infeasible solution. As we increase the threshold, projects are removed in sequence (without adding new projects), which monotonically decreases the maximum violation. For a general boolean LP, the corresponding plots need not exhibit monotonic behavior.

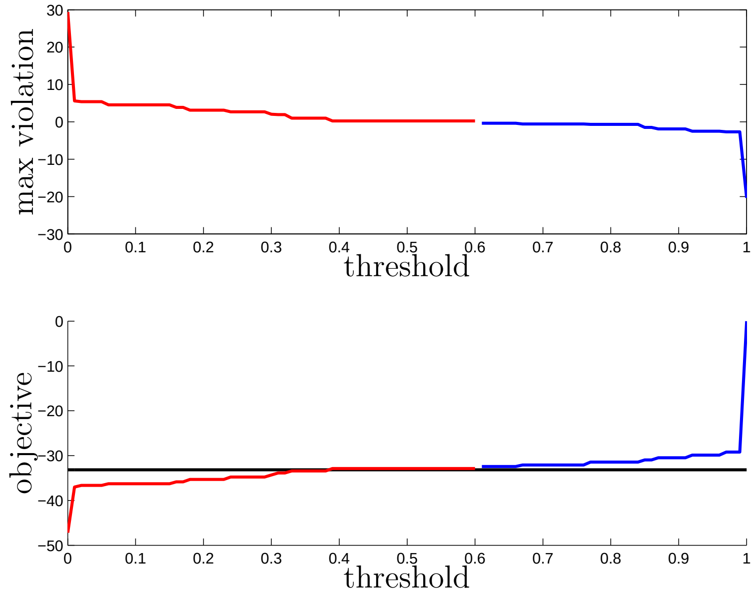


Figure 1 Plots of violation and objective vs threshold rule.

3. *Maximizing house profit in a gamble and imputed probabilities.* A set of n participants bet on which one of m outcomes, labeled $1, \ldots, m$, will occur. Participant i offers to purchase up to $q_i > 0$ gambling contracts, at price $p_i > 0$, that the true outcome will be in the set $S_i \subset \{1, \ldots, m\}$. The house then sells her x_i contracts, with $0 \leq x_i \leq q_i$. If the true outcome j is in S_i, then participant i receives \$1 per contract, *i.e.*, x_i. Otherwise, she loses, and receives nothing. The house collects a total of $x_1 p_1 + \cdots + x_n p_n$, and pays out an amount that depends on the outcome j,

$$\sum_{j \in S_i} x_i.$$

The difference is the house profit.

   (a) *Optimal house strategy.* How should the house decide on x so that its worst-case profit (over the possible outcomes) is maximized? (The house determines x after examining all the participant offers.)

   (b) *Imputed probabilities.* Suppose $x^\star$ maximizes the worst-case house profit. Show that there exists a probability distribution π on the possible outcomes (*i.e.*, $\pi \in \mathbf{R}_+^m$, $\mathbf{1}^T \pi = 1$) for which $x^\star$ also maximizes the expected house profit. Explain how to find π.

   Hint. Formulate the problem in part (a) as an LP; you can construct π from optimal dual variables for this LP.

   Remark. Given π, the 'fair' price for offer i is $p_i^{\text{fair}} = \sum_{j \in S_i} \pi_j$. All offers with $p_i > p_i^{\text{fair}}$ will be completely filled (*i.e.*, $x_i = q_i$); all offers with $p_i < p_i^{\text{fair}}$ will be rejected (*i.e.*, $x_i = 0$).

Remark. This exercise shows how the probabilities of outcomes (*e.g.*, elections) can be guessed from the offers of a set of gamblers.

(c) *Numerical example.* Carry out your method on the simple example below with $n = 5$ participants, $m = 5$ possible outcomes, and participant offers

Participant i	p_i	q_i	S_i
1	0.50	10	{1,2}
2	0.60	5	{4}
3	0.60	5	{1,4,5}
4	0.60	20	{2,5}
5	0.20	10	{3}

Compare the optimal worst-case house profit with the worst-case house profit, if all offers were accepted (*i.e.*, $x_i = q_i$). Find the imputed probabilities.

Solution.

(a) The worst-case house profit is

$$p^T x - \max_{j=1,\ldots,m} \sum_{j \in S_i} x_i,$$

which is a piecewise-linear concave function of x. To find the x that maximizes the worst-case profit, we solve the problem,

$$\begin{array}{ll} \text{maximize} & p^T x - \max_{j=1,\ldots,m} a_j^T x \\ \text{subject to} & 0 \preceq x \preceq q, \end{array}$$

with variable x. a_j^T are the rows of the subset matrix A, with

$$A_{ji} = \begin{cases} 1 & j \in S_i \\ 0 & \text{otherwise.} \end{cases}$$

(b) The problem from part (a) can be expressed as

$$\begin{array}{ll} \text{maximize} & p^T x - t \\ \text{subject to} & t\mathbf{1} \succeq Ax \\ & 0 \preceq x \preceq q, \end{array} \tag{2}$$

where t is a new scalar variable. The Lagrangian is

$$L(x, t, \lambda_1, \lambda_2, \lambda_3) = t - p^T x + \lambda_1^T (Ax - t\mathbf{1}) - \lambda_2^T x + \lambda_3^T (x - q).$$

This is bounded below if and only if $\mathbf{1}^T \lambda_1 = 1$, and $A^T \lambda_1 - \lambda_2 + \lambda_3 = p$. The dual can be written as

$$\begin{array}{ll} \text{maximize} & -q^T \lambda_3 \\ \text{subject to} & \mathbf{1}^T \lambda_1 = 1 \\ & A^T \lambda_1 - \lambda_2 + \lambda_3 = p \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0, \quad \lambda_3 \succeq 0, \end{array} \tag{3}$$

with variables λ_1, λ_2, and λ_3. Notice that λ_1 must satisfy $\mathbf{1}^T\lambda_1 = 1$, and $\lambda_1 \succeq 0$, hence it is a probability distribution.

Suppose $x^\star_{\rm wc}$, $t^\star$, $\lambda^\star_1$, $\lambda^\star_2$, and $\lambda^\star_3$ are primal and dual optimal for problem (2), and let us set $\pi = \lambda^\star_1$. To maximize the expected house profit we solve the problem,

$$\begin{array}{ll} \mbox{maximize} & p^T x - \pi^T A x \\ \mbox{subject to} & 0 \preceq x \preceq q. \end{array} \qquad (4)$$

Let $b_1^T, \ldots, b_n^T$ be the rows of A^T. We know that a point $x^\star_{\rm e}$ is optimal for problem (4) if and only if $x^\star_{{\rm e}i} = q_i$ when $p_i - b_i^T\pi > 0$, $x^\star_{{\rm e}i} = 0$ when $p_i - b_i^T\pi < 0$, and $0 \le x^\star_{{\rm e}i} \le q_i$ when $p_i - b_i^T\pi = 0$.

To see why $x^\star_{\rm e} = x^\star_{\rm wc}$, let us take a look at one of the KKT conditions for problem (2). This can be written as

$$p - A^T\pi = \lambda^\star_3 - \lambda^\star_2,$$

with $\pi = \lambda^\star_1$. If $p_i - b_i^T\pi > 0$, then we must have $\lambda^\star_{3i} - \lambda^\star_{2i} > 0$, which means that $\lambda^\star_{2i} = 0$ and $\lambda^\star_{3i} > 0$ (by complementary slackness), and so $x^\star_{{\rm wc}i} = q_i$. Similarly, if $p_i - b_i^T\pi < 0$, then $\lambda^\star_{3i} - \lambda^\star_{2i} < 0$, which means that $\lambda^\star_{2i} > 0$ and $\lambda^\star_{3i} = 0$, and so $x^\star_{{\rm wc}i} = 0$. Finally, when $p_i - b_i^T\pi = 0$, we must have $\lambda^\star_{2i} = 0$ and $\lambda^\star_{3i} = 0$, and so $0 \le x^\star_{{\rm wc}i} \le q_i$.

In summary, in order to find a probability distribution on the possible outcomes for which the same $x^\star$ maximizes both the worst-case as well as the expected house profit, we solve the dual LP (3), and set $\pi = \lambda^\star_1$.

(c) The following `cvx` code solves the problem.

```
% solution for gambling problem
A = [1 0 1 0 0;
     1 0 0 1 0;
     0 0 0 0 1;
     0 1 1 0 0;
     0 0 1 1 0];

p = [0.5; 0.6; 0.6; 0.6; 0.2];
q = [10; 5; 5; 20; 10];

n = 5; m = 5;

cvx_begin
    variables x(n) t
    dual variable lambda1
    maximize (p'*x-t)
    subject to
```

```
        lambda1: A*x <= t
        x >= 0
        x <= q
cvx_end

% optimal worst case house profit
pwc = cvx_optval
% optimal worst case profit if all offer are accepted
pwc_accept = p'*q-max(A*q)
% imputed probabilities
pi = lambda1
% fair prices
pfair = A'*pi
% optimal purchase quantities
xopt = x
```

Our results are summarized in the following table. We find that the optimal worst case house profit is 3.5, and the worst case house profit, if all offers are accepted, is -5. The imputed probabilities are

$$\pi = (0.1145,\ 0.3855,\ 0.0945,\ 0.1910,\ 0.2145).$$

The associated fair prices and optimal contract numbers are shown below.

Participant i	p_i	p_i^{fair}	q_i	x_i	S_i
1	0.50	0.5000	10	5	{1,2}
2	0.60	0.1910	5	5	{4}
3	0.60	0.5200	5	5	{1,4,5}
4	0.60	0.6000	20	5	{2,5}
5	0.20	0.0945	10	10	{3}

4. *Minimax rational fit to the exponential.* (See exercise 6.9.) We consider the specific problem instance with data

$$t_i = -3 + 6(i-1)/(k-1), \quad y_i = e^{t_i}, \quad i = 1, \ldots, k,$$

where $k = 201$. (In other words, the data are obtained by uniformly sampling the exponential function over the interval $[-3, 3]$.) Find a function of the form

$$f(t) = \frac{a_0 + a_1 t + a_2 t^2}{1 + b_1 t + b_2 t^2}$$

that minimizes $\max_{i=1,\ldots,k} |f(t_i) - y_i|$. (We require that $1 + b_1 t_i + b_2 t_i^2 > 0$ for $i = 1, \ldots, k$.)

Find optimal values of $a_0,\ a_1,\ a_2,\ b_1,\ b_2$, and give the optimal objective value, computed to an accuracy of 0.001. Plot the data and the optimal rational function fit on the same plot. On a different plot, give the fitting error, *i.e.*, $f(t_i) - y_i$.

Hint. You can use `strcmp(cvx_status,'Solved')`, after `cvx_end`, to check if a feasibility problem is feasible.

Solution. The objective function (and therefore also the problem) is not convex, but it is quasiconvex. We have $\max_{i=1,\dots,k} |f(t_i) - y_i| \leq \gamma$ if and only if

$$\left| \frac{a_0 + a_1 t_i + a_2 t_i^2}{1 + b_1 t_i + b_2 t_i^2} - y_i \right| \leq \gamma, \quad i = 1, \dots, k.$$

This is equivalent to (since the denominator is positive)

$$|a_0 + a_1 t + a_2 t_i^2 - y_i(1 + b_1 t_i + b_2 t_i^2)| \leq \gamma(1 + b_1 t_i + b_2 t_i^2), \quad i = 1, \dots, k,$$

which is a set of $2k$ linear inequalities in the variables a and b (for fixed γ). In particular, this shows the objective is quasiconvex. (In fact, it is a generalized linear fractional function.)

To solve the problem we can use a bisection method, solving an LP feasibility problem at each step. At each step we select some value of γ and solve the feasibility problem

$$\begin{array}{ll} \text{find} & a,\ b \\ \text{subject to} & |a_0 + a_1 t_i + a_2 t_i^2 - y_i(1 + b_1 t_i + b_2 t_i^2)| \leq \gamma(1 + b_1 t_i + b_2 t_i^2), \quad i = 1, \dots, k, \end{array}$$

with variables a and b. (Note that as long as $\gamma > 0$, the condition that the denominator is positive is enforced automatically.) This can be turned into the LP feasibility problem

$$\begin{array}{ll} \text{find} & a,\ b \\ \text{subject to} & a_0 + a_1 t_i + a_2 t_i^2 - y_i(1 + b_1 t_i + b_2 t_i^2) \leq \gamma(1 + b_1 t_i + b_2 t_i^2), \quad i = 1, \dots, k \\ & a_0 + a_1 t_i + a_2 t_i^2 - y_i(1 + b_1 t_i + b_2 t_i^2) \geq -\gamma(1 + b_1 t_i + b_2 t_i^2), \quad i = 1, \dots, k. \end{array}$$

The following Matlab code solves the problem for the particular problem instance.

```
k=201;
t=(-3:6/(k-1):3)';
y=exp(t);

Tpowers=[ones(k,1) t t.^2];

u=exp(3); l=0; % initial upper and lower bounds
bisection_tol=1e-3; % bisection tolerance

while u-l>= bisection_tol
    gamma=(l+u)/2;
    cvx_begin % solve the feasibility problem
    cvx_quiet(true);
```

```
    variable a(3);
    variable b(2);
    subject to
        abs(Tpowers*a-y.*(Tpowers*[1;b])) <= gamma*Tpowers*[1;b];
    cvx_end

    if strcmp(cvx_status,'Solved')
        u=gamma;
        a_opt=a;
        b_opt=b;
objval_opt=gamma;
    else
        l=gamma;
    end
end

y_fit=Tpowers*a_opt./(Tpowers*[1;b_opt]);

figure(1);
plot(t,y,'b', t,y_fit,'r+');
xlabel('t');
ylabel('y');

figure(2);
plot(t, y_fit-y);
xlabel('t');
ylabel('err');
```

The optimal values are

$$a_0 = 1.0099, \quad a_1 = 0.6117, \quad a_2 = 0.1134, \quad b_1 = -0.4147, \quad b_2 = 0.0485,$$

and the optimal objective value is 0.0233. We also get the following plots.

5. *Maximum likelihood estimation of x and noise mean and covariance.* Consider the maximum likelihood estimation problem with the linear measurement model

$$y_i = a_i^T x + v_i, \quad i = 1, \dots, m$$

(as discussed on page 7-3 of the lecture notes and page 352 of the textbook). The vector $x \in \mathbf{R}^n$ is a vector of unknown parameters, y_i are the measurement values, and v_i are independent and identically distributed measurement errors.

In this problem we make the assumption that the *normalized* probability density function of the errors is given (normalized to have zero mean and unit variance), but not

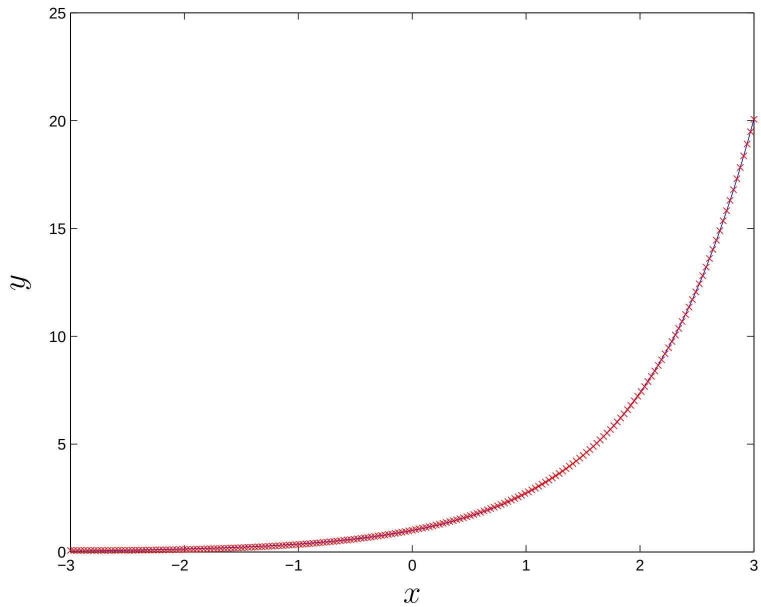


Figure 2 Chebyshev fit with rational function. The line represents the data and the crosses the fitted points.

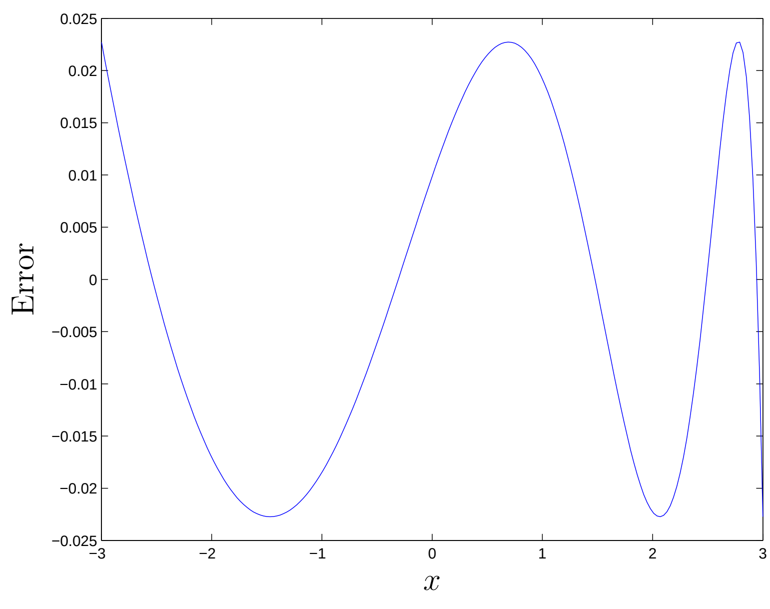


Figure 3 Fitting error for Chebyshev fit of exponential with rational function.

their mean and variance. In other words, the density of the measurement errors v_i is

$$p(z) = \frac{1}{\sigma} f(\frac{z-\mu}{\sigma}),$$

where f is a given, normalized density. The parameters μ and σ are the mean and standard deviation of the distribution p, and are not known.

The maximum likelihood estimates of x, μ, σ are the maximizers of the log-likelihood function

$$\sum_{i=1}^{m} \log p(y_i - a_i^T x) = -m \log \sigma + \sum_{i=1}^{m} \log f(\frac{y_i - a_i^T x - \mu}{\sigma}),$$

where y is the observed value. Show that if f is log-concave, then the maximum likelihood estimates of x, μ, σ can be determined by solving a convex optimization problem.

Solution. With a change of variables

$$z = (1/\sigma)x, \qquad u = \mu/\sigma, \qquad t = 1/\sigma,$$

the problem reduces to

$$\text{maximize} \quad m \log t + \sum_{i=1}^{m} \log f(y_i t - a_i^T z - u),$$

which is a convex optimization problem since the objectve is a concave function of (z, u, t). We recover optimal values of x, μ, and σ using

$$\sigma = 1/t, \qquad \mu = u/t, \qquad x = (1/t)z.$$

6. *Maximum likelihood estimation for exponential family.* A probability distribution on $\mathcal{D} \subseteq \mathbf{R}^n$, parametrized by $\theta \in \mathbf{R}^m$, is called an *exponential family* if it has the form

$$p_\theta(x) = a(\theta) \exp(\theta^T c(x))$$

for $x \in \mathcal{D}$, where $c : \mathbf{R}^n \to \mathbf{R}^m$, and

$$a(\theta) = \left(\int_{\mathcal{D}} \exp(\theta^T c(x)) \, dx \right)^{-1}.$$

(We consider only values of θ for which the integral above is finite.) Many families of distributions have this form, for appropriate choice of the parameter θ.

(a) When $c(x) = x$ and $\mathcal{D} = \mathbf{R}_+^n$, what is the associated family of distributions? What is the set of valid values of θ?

(b) Explain how to represent the normal family $\mathcal{N}(\mu, \Sigma)$ as an exponential family. *Hint.* Use parameter $(z, Y) = (\Sigma^{-1}\mu, \Sigma^{-1})$. With this parameter, $\theta^T c(x)$ has the form $z^T c_1(x) + \mathbf{tr}\, Y C_2(x)$, where $C_2(x) \in \mathbf{S}^n$.

(c) Show that for any $x \in \mathcal{D}$, the log-likelihood function $\log p_\theta(x)$ is concave in θ. This means that maximum-likelihood estimation for an exponential family leads to a convex optimization problem. You don't have to give a formal proof of concavity of $\log p_\theta(x)$; if you like, you can approximate the integral appearing in the expression as a (finite) Riemann sum, show concavity of this approxmation, and then just state 'take the limit'.

Solution.

(a) We need $\theta \prec 0$ for the integral to converge, in which case we have

$$\int_{\mathbf{R}^n_+} \exp(\theta^T x)\, dx = \frac{1}{\prod_{i=1}^n (-\theta_i)},$$

so $a(\theta) = \prod_{i=1}^n (-\theta_i)$. In this case the distribution is that of n independent exponentially distributed variables, with means $-1/\theta_i$.

(b) We write the $\mathcal{N}(\mu, \Sigma)$ density as

$$p(x) = a \exp\left(-(x-\mu)^T \Sigma^{-1}(x-\mu)/2\right),$$

where a is the normalizing constant (which depends on μ and Σ). We write this in the form

$$p(x) = \tilde{a} \exp\left(z^T c_1(x) + \mathbf{tr}\, Y C_2(x)\right),$$

where $Y = \Sigma^{-1}$ and $z = \Sigma^{-1}\mu$,

$$c_1(x) = x, \qquad C_2(x) = -(1/2)xx^T,$$

and $\tilde{a}$ is the normalizing constant (which depends on Y and z). The set of valid values of z and Y is simply $\mathbf{R}^n \times \mathbf{S}^n_{++}$. (Note that the mapping between (μ, Σ) and (z, Y) is a bijection.)

(c) From the definition of $p_\theta(x)$ we have

$$\begin{aligned} \log p_\theta(x) &= \log a(\theta) + \theta^T c(x) \\ &= -\log\left(\int_{\mathcal{D}} \exp(\theta^T c(x))\, dx\right) + \theta^T c(x). \end{aligned}$$

The second term is affine in θ; the first is concave, but this is harder to show. We'll follow the hint and approximate the integral as a finite sum, by approximating $c(x)$ as having values c_i on N regions, with volumes A_i:

$$-\log\left(\int_{\mathcal{D}} \exp(\theta^T c(x))\, dx\right) \approx -\log\left(\sum_{i=1}^N A_i \exp(\theta^T c_i)\right).$$

The righthand side is evidently concave, since it is minus the log-sum-exp function composed with the affine function of θ given by

$$F(\theta)_i = c_i^T\theta + \log A_i, \quad i = 1, \ldots, N.$$

Now we just say (with hands waving) 'Take the limit'.

We can also show that $f(\theta) = \log p_\theta(x)$ is concave by computing the Hessian. Defining $e(\theta, x) = \exp(\theta^T c(x))$, the second partial derivatives are

$$\frac{\partial^2 f}{\partial\theta_i\partial\theta_j} = \frac{\left(\int_{\mathcal{D}} c_i(x)e(\theta,x)\; dx\right)\left(\int_{\mathcal{D}} c_j(x)e(\theta,x)\; dx\right) - \left(\int_{\mathcal{D}} c_i(x)c_j(x)e(\theta,x)\; dx\right)\left(\int_{\mathcal{D}} e(\theta,x)\; dx\right)}{\left(\int_{\mathcal{D}} e(\theta,x)\; dx\right)^2},$$

where $c(x) = (c_1(x), \ldots, c_m(x))$. For any vector $v \in \mathbf{R}^m$, this gives

$$v^T(\nabla^2 f(\theta))v = \frac{\left(\int_{\mathcal{D}}(\sum_i v_i c_i(x))e(\theta,x)\; dx\right)^2 - \left(\int_{\mathcal{D}}(\sum_i v_i c_i(x))^2 e(\theta,x)\; dx\right)\left(\int_{\mathcal{D}} e(\theta,x)\; dx\right)}{\left(\int_{\mathcal{D}} e(\theta,x)\; dx\right)^2}.$$

From the Cauchy-Schwarz inequality for functions,

$$\left(\int g(x)h(x)\; dx\right)^2 \le \left(\int (g(x))^2\; dx\right)\left(\int (h(x))^2\; dx\right),$$

with $g(x) = (\sum_i v_i c_i(x))\sqrt{e(\theta,x)}$ and $h(x) = \sqrt{e(\theta,x)}$, we get $v^T(\nabla^2 f(\theta))v \le 0$ for all v. The Hessian is negative semidefinite, so $\log p_\theta(x)$ is concave.