10pts

First we show that sum of two self-concordant function is self-concordant. Note that $f''(x) \geq 0$ for all $x \in (a, b)$ if $f$ is convex in $(a, b)$.

$$|f_1''(x) + f_2''(x)| \leq |f_1''(x)| + |f_2''(x)|$$
$$\leq 2(|f_1''(x)|^{3/2} + |f_2''(x)|^{3/2})$$
$$\leq 2(|f_1''(x) + f_2''(x)|)^{3/2},$$

where we use the following inequality in the last step.

$$(u^{3/2} + v^{3/2})^{2/3} \leq u + v,$$

which holds for $u, v \geq 0$.

1. If both $f$ and $g$ are convex functions and $\lim_{x \to D} f(x) = \infty$, $\lim_{x \to D} g(x) = \infty$ then $\lim_{x \to D} f(x) + g(x) = \infty$. Note that it follows from convexity of $f, g$ that they are finite on the interior of their domains.

We showed above that sum of two self-concordant functions is self-concordant above. Here is the proof of the third property of self-concordant functions.

$$|(f + g)'_{x,u}(t)| = |(f + g)'(x + tu)|$$
$$= |f'(x + tu) + g'(x + tu)|$$
$$= |f'_{x,u}(t) + g'_{x,u}(t)|$$
$$\leq |f'_{x,u}(t)| + |g'_{x,u}(t)|$$
$$\leq (\alpha |f''_{x,u}(t)|)^{1/2} + (\beta |g''_{x,u}(t)|)^{1/2}$$
$$\leq \left( (\alpha + \beta) |f''_{x,u}(t) + g''_{x,u}(t)| \right)^{1/2},$$

where the last step follows from the fact $(a - b)^2 \geq 0$. 


2. Let \( f(x_i) = -\ln(x_i) \)
   - \( \lim_{x_i \to 0^+} -\ln(x_i) = \infty \)
   - \( f'(x_i) = \frac{-1}{x_i}, \quad f''(x_i) = \frac{1}{x_i^2} \) and \( f'''(x_i) = \frac{-2}{x_i^3} \). Thus, \( |f'''(x_i)| \leq 2|f''(x_i)|^{3/2} \) holds.
   - By above calculations, it is straightforward to check that this condition holds too.

3. Follows from (a) and (b).
2 10pts

1. Suppose for contradiction that an extreme point solution $x$ is not half-integral. Let $J = \{x_i | x_i \notin \{0, 1/2, 1\}\}$ and let $\epsilon = \min_{i \in J} \min \{x_i, |1/2 - x_i|, |1 - x_i|\}$. Define the following vectors $x^1, x^2$ derived from $x$:

\[
x^1_i = \begin{cases} 
  x_i & \text{if } x_i \in \{0, 1/2, 1\} \\
  x_i - \epsilon & \text{if } x_i < 1/2 \\
  x_i + \epsilon & \text{if } x_i > 1/2 
\end{cases}
\]

and,

\[
x^2_i = \begin{cases} 
  x_i & \text{if } x_i \in \{0, 1/2, 1\} \\
  x_i + \epsilon & \text{if } x_i < 1/2 \\
  x_i - \epsilon & \text{if } x_i > 1/2 
\end{cases}
\]

It is easy to check that $x = \frac{1}{2}(x^1 + x^2)$ and $x \neq x^1 \neq x^2$ which contradicts the fact that $x$ is an extreme point solution.

2. Let $x$ be an optimal extreme point solution of the LP relaxation of Vertex Cover. By part (a), $x$ is a half-integral solution. Let $S = \{i | x_i = 1\}$ and $T = \{i | x_i = 1/2\}$. Since the underlying graph is planar, it is 4-colorable. Let $T_j$ be the set of vertices in $T$ of color $j$ where $j$ is one of the four colors. $(S \cup T) - \arg\min_{j} w(T_j)$ is a $3/2$-approximate solution.
For each clause $C$ we define $P$ to contain the indices of all variable that occur positively in $C$ and $N$ to be the set of indices that occur negatively in $C$; in other words, each clause $C_j$ can be written as

\[ \bigvee_{i \in P_j} x_i \land \bigwedge_{i \in N_j} \bar{x}_i. \]

The constraint corresponds to each clause is the following,

\[ \sum_{i \in P_j} y_i + \sum_{i \in N_j} (1 - y_i) \geq z_j. \]

The LP-relaxation of MAX-SAT problem is the following.

\[
\begin{align*}
\max & \quad \sum_{j=1}^{m} w_j z_j \\
\text{s.t.} & \quad \sum_{i \in P_j} y_i + \sum_{i \in N_j} (1 - y_i) \geq z_j \quad \forall j \leq m \\
& \quad 0 \leq y_i \leq 1 \quad \forall i \leq n \\
& \quad 0 \leq z_j \leq 1 \quad \forall j \leq m
\end{align*}
\]

2. Let $(y^*, z^*)$ be an optimal solution of the above LP-relaxation. We now analyze the performance of the algorithm that sets each $x_i$ true with probability $y_i$.

\[
\Pr[\text{clause } C_j \text{ is not satisfied}] = \prod_{i \in P_j} (1 - y_i^*) \prod_{i \in N_j} (y_i^*) \leq \left( \frac{1}{\ell_j} \left( \sum_{i \in P_j} (1 - y_i^*) + \sum_{i \in N_j} y_i^* \right) \right)^{\ell_j},
\]

where the last step is by arithmetic-geometric mean inequality and $\ell_j$ denote the number of literals is $C_j$. Note that

\[
\left( \frac{1}{\ell_j} \left( \sum_{i \in P_j} (1 - y_i^*) + \sum_{i \in N_j} y_i^* \right) \right)^{\ell_j} = \left( 1 - \frac{1}{\ell_j} \left( \sum_{i \in P_j} y_i^* + \sum_{i \in N_j} (1 - y_i^*) \right) \right)^{\ell_j}.
\]

Using the fact that $(y^*, z^*)$ is a (optimal) feasible solution, i.e. $\sum_{i \in P_j} y_i + \sum_{i \in N_j} (1 - y_i) \geq z_j$,

\[
\Pr[\text{clause } C_j \text{ is not satisfied}] \leq \left( 1 - \frac{z_j^*}{\ell_j} \right)^{\ell_j}
\]
By Concavity of \((1 - \frac{z_j^*}{\ell_j})^{\ell_j}\),

\[
\Pr[\text{clause } C_j \text{ is satisfied}] \leq \left[ 1 - \left( 1 - \frac{1}{\ell_j} \right)^{\ell_j} \right] z_j^*
\]

Hence the expected number of satisfied clauses is

\[
\sum_{j=1}^{m} w_j \cdot \Pr[\text{clause } C_j \text{ is satisfied}] = \sum_{j=1}^{m} \left[ 1 - \left( 1 - \frac{1}{\ell_j} \right)^{\ell_j} \right] z_j^*
\]

\[
\leq \min_{j \leq m} \left[ 1 - \left( 1 - \frac{1}{\ell_j} \right)^{\ell_j} \right] \sum_{j \leq m} w_j z_j^*
\]

\[
\leq (1 - \frac{1}{e}) \text{OPT}.
\]

3. The unbiased randomized algorithm satisfies each clause \(C_j\) with probability \(1 - 2^{-\ell_j}\).

In order to compute the performance of the maximum of \(W_1\) (randomized rounding approach) and \(W_2\) (unbiased randomized algorithm), we can compute \(\frac{1}{2}(W_1 + W_2)\).

\[
\frac{1}{2}(W_1 + W_2) \geq \frac{1}{2} \sum_{j \leq m} w_j z_j^* \left[ 1 - \left( 1 - \frac{1}{\ell_j} \right)^{\ell_j} \right] + (1 - 2^{-\ell_j}) \geq \frac{3}{4} \sum_{j \leq m} w_j z_j^*
\]

where the last step can be proved via a case by case analysis of \(\ell_j\) (\(\ell_j = 1, \ell_j = 2\) and \(\ell_j \geq 3\)).
4 10pts

(a)
You can do this by just showing that the Hessian is $2A^TA$, which is PSD. Alternatively, we can show directly that for any $\lambda \in [0,1]$,
\[
\lambda f(x) + (1 - \lambda) f(y) \geq f(x + (1 - \lambda)y)
\]
Rewrite $f(x) = x^TA^T Ax + 2x^TA^T b + b^T b$. Then:
\[
f(\lambda x + (1 - \lambda)y) = \lambda^2 x^T A^T Ax + (1 - \lambda)^2 y^T A^T Ay + 2\lambda(1 - \lambda)x^T A^T Ay
+ 2(\lambda x + (1 - \lambda)y)A^T b + b^T b
\]
Considering just the terms on the first line:
\[
\begin{align*}
\lambda^2 x^T A^T Ax + (1 - \lambda)^2 y^T A^T Ay &+ 2\lambda(1 - \lambda)x^T A^T Ay \\
\leq \lambda^2 x^T A^T Ax + (1 - \lambda)^2 y^T A^T Ay &+ 2\lambda(1 - \lambda)\sqrt{x^T A^T Ax \cdot y^T A^T Ay} \\
\leq \lambda^2 x^T A^T Ax + (1 - \lambda)^2 y^T A^T Ay &+ \lambda(1 - \lambda)(x^T A^T Ax + y^T A^T Ay) \\
= \lambda x^T A^T Ax + (1 - \lambda)y^T A^T Ay
\end{align*}
\]
The inequalities follow from Cauchy-Schwarz and AM-GM inequality. Combining with the second line unchanged gives the convexity inequality.

(b)
Since $f(x) = x^T A^T Ax - 2x^T A^T b + b^T b$, $\nabla f(x) = 2A^T Ax - 2A^T b$. So, the iteration we’re looking at is just gradient descent with learning rate $\eta$.

As shown in class, for convergence we need that $\eta \leq 1/\beta$ where $\beta$ is the smooth convexity of our function. Let’s check the smooth convexity of our function
\[
\|\nabla f(x) - \nabla f(y)\|_2 = \|2A^T A(x - y)\|_2 \leq 2\|A^T A\|_2 \|(x - y)\|_2
\]
So $\beta = 2\|A^T A\|_2$ and therefore setting $\eta$ as described is sufficient.

To check that our iteration converges, we’re also going to need to compute the strong-convexity of $f(x)$. Note that $\nabla^2 f(x) = 2A^T A$. Assuming $A$ is full rank,
\[
z^T (2A^T A) z \geq \frac{2}{\|(A^T A)^{-1}\|_2} = \alpha
\]
In class we showed that after $t$ iterations, we can guarantee:
\[
f(x) - f(x^*) \leq \beta (1 - \eta \alpha / 2)^{t-1} \|x^* - x_0\|_2.
\]
Now $\|x^* - x_0\|_2 = \|(A^T A)^{-1} A^T b - 0\|_2 \leq \|(A^T A)^{-1}\|_2 \|A^T b\|_2 \leq \|(A^T A)^{-1}\|_2 \|A^T A\|_2$.

Substituting $\beta = \|A^T A\|_2$, to get our guarantee we need:
\[
(1 - \eta \alpha / 2)^{t-1} \leq \epsilon \|(A^T A)^{-1}\|_2 \|A^T A\|_2
\]
\[
(1 - \eta \alpha / 2) = (1 - O(\frac{1}{\|(A^T A)^{-1}\|_2 \|A^T A\|_2}))
\]
so we need $t = \|(A^T A)^{-1}\|_2 \|A^T A\|_2$ to get the expression down to a constant and then just log terms for the error we want.
(c)

\[ x_{t+1} = Mx_t + 2\eta A^T b \]
\[ = Mx_t + 2\eta A^T Ax^* \]
\[ = M(x_t - x^*) + x^* \]

(d)

So our error evolves as \((x_{t+1} - x^*) \leq M(x_t - x^*).\) Accordingly, we simply need \(\|M\|_2 < 1\) to guarantee convergence. Suppose we set \(\eta = \frac{1}{2\|A^T A\|_2}.\) Then we have guarantee that all of the eigenvalues of \(A\) lie between 0 and 1. Specifically, the largest will have value \(1 - \frac{1}{2\|A^T A\|_2\|A^T A^{-1}\|_2}.\) Accordingly, we need to power up \(M O(\|A^T A\|_2\|A^T A^{-1}\|_2 \log(1/\epsilon))\) times to get error \(\epsilon.\)