Problem 1 50 points

(a) Alternative (exact Hessian) 25 points

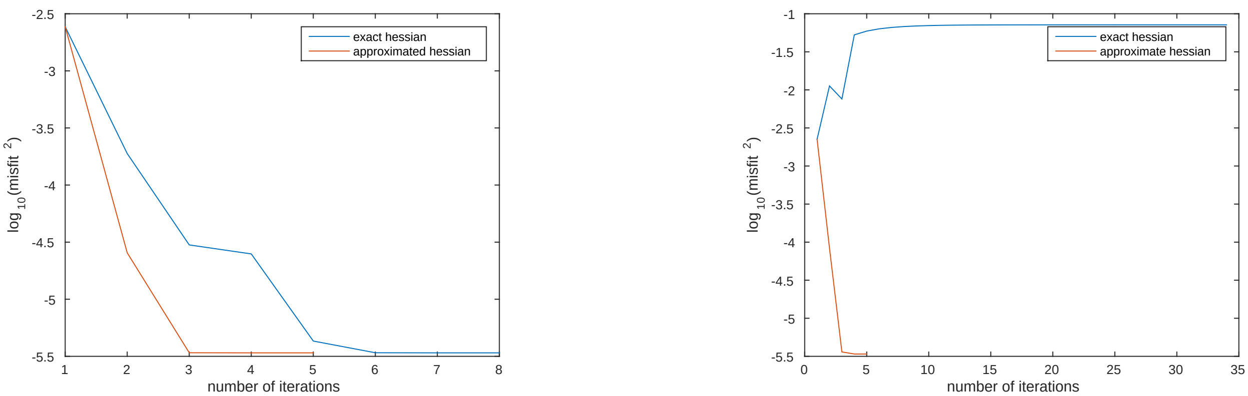

Figure 1: Left: In one example where initial guess is not very close to solution, using exact Hessian takes more iterations than using approximate Hessian. Right: In another example, using exact Hessian donot converge while using approximate Hessian converges.

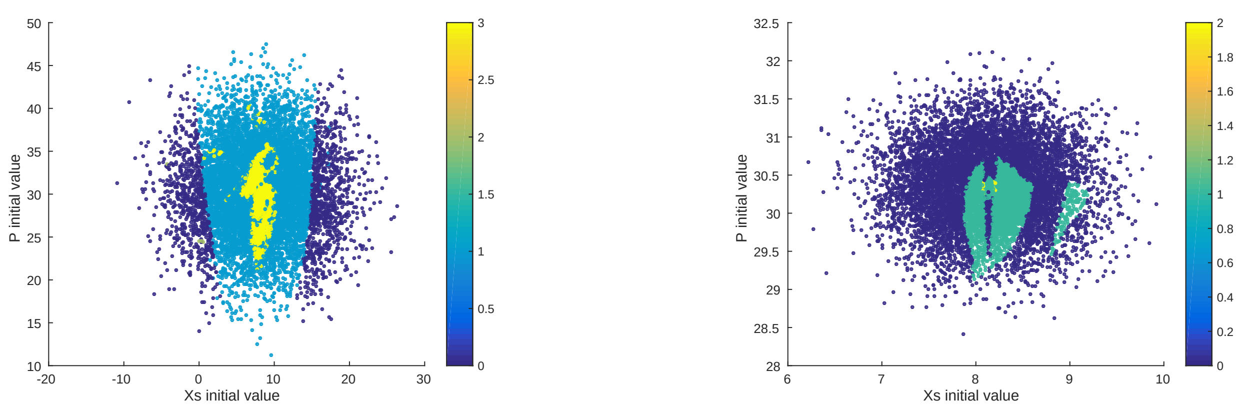

Figure 2: Left:Convergence comparison of approximate Hessian and exact one. Dark blue region is where both methods donot converge. Light blue region is where approximate Hessian converge but exact Hessian don't. Yellow region is where both methods converge. Right:Dark blue region is where approximate Hessian takes less iterations.

The conclusion is that when using exact Hessian, the initial guess has to be closer to the solution so as to make the iteration converge. If the initial guess is not very close to the solution, approximate Hessian will be more robust which means both a better chance of finding the optimal solution and less iterations. Only when initial guess is very close to solution, exact Hessian will take equal or less iterations than approximate Hessian.

```
function [ Gamma, Hess] = compute_gradient_approx_hess2( x,y,M, misfit)

xs = M(1);
ys = M(2);
zs = M(3);
p = M(4);

R = ((x - xs).^2 + (y - ys).^2 + zs^2);

dx = x-xs;
dy = y-ys;

Ghat(:,1) = (3.*p.*zs.*(dx))./((R).^(5/2));
Ghat(:,2) = (3.*p.*zs.*(dy))./((R).^(5/2));
Ghat(:,3) = p./(R).^(3/2) - (3*p.*zs.^2)./(R).^(5/2);
Ghat(:,4) = zs./(R).^(3/2);

Gamma = - Ghat'*misfit;
%this is the apprximated Hessian;


dHess = zeros(4);
dHess(1,2) = - misfit'*((15*p*zs*(dx).*(dy))./((R).^(7/2)));
dHess(1,3) = - misfit'*((3*p.*(dx))./((R).^(5/2)) - (15*p*zs.^2.*(dx))./((R).^(7/2)));
dHess(1,4) = - misfit'*((3*zs.*(dx))./((R).^(5/2)));
dHess(2,3) = - misfit'*( (3*p.*(dy))./((R).^(5/2)) - (15*p*zs.^2.*(dy))./((R).^(7/2)));
dHess(2,4) = - misfit'*((3*zs.*(dy))./((R).^(5/2)));
dHess(3,4) = - misfit'*(1./(R).^(3/2) - (3*zs^2)./(R).^(5/2));

dHess = (dHess + dHess');

dHess(1,1) = - misfit'*((15*p*zs.*(dx).^2)./((R).^(7/2)) - (3*p.*zs)./(R).^(5/2));
dHess(2,2) = - misfit'*((15*p*zs.*(dy).^2)./((R).^(7/2)) - (3*p.*zs)./(R).^(5/2));
dHess(3,3) = - misfit'*((15*p*zs^3)./(R).^(7/2) - (9*p*zs)./(R).^(5/2));
dHess(4,4) = 0;



Hess = (Ghat')*Ghat;
%only add dHess if using exact Hessian
Hess = Hess + dHess;


end
```

(b) 25 points

(b.i)

Change variable $t = 1/\sigma$, then

$$\int_0^\infty \frac{1}{\sigma^N} \exp\left(-\frac{1}{2\sigma^2}A\right) d\sigma \quad = \quad \int_0^\infty t^{N-2} \exp\left(-\frac{A}{2}t^2\right) dt$$

Since

$$\int_0^\infty \exp\left(-\frac{A}{2}t^2\right) dt = \sqrt{\frac{\pi}{2}} A^{-1/2} \propto A^{-1/2}$$

$$\int_0^\infty t \exp\left(-\frac{A}{2}t^2\right) dt = A^{-1} \propto A^{-1}$$

Take derivative with respect to A, then

$$\int_0^\infty \frac{-t^2}{2} \exp\left(-\frac{A}{2}t^2\right) dt \propto A^{-1/2-1}$$

$$\int_0^\infty t\frac{-t^2}{2} \exp\left(-\frac{A}{2}t^2\right) dt \propto A^{-1-1}$$

Continue this derivative, we have

$$\int_0^\infty \left(\frac{-t^2}{2}\right)^{(N-2)/2} \exp\left(-\frac{A}{2}t^2\right) dt \propto A^{-1/2-(N-2)/2}$$

So

$$\int_0^\infty t^{N-2} \exp\left(-\frac{A}{2}t^2\right) dt \propto A^{-(N-1)/2}$$

(b.ii)

Assume uniform prior, then

$$\begin{aligned}P(x,y|d) &\propto P(d|x,y) = \exp\left(-F(x,y)\right)\\ &= \exp\left(-F(x_0,y_0) - \nabla F|_{x_0,y_0}\cdot\begin{bmatrix} x-x_0 & y-y_0 \end{bmatrix}' - \frac{1}{2}\begin{bmatrix} x-x_0 & y-y_0 \end{bmatrix} H|_{x_0,y_0}\begin{bmatrix} x-x_0 & y-y_0 \end{bmatrix}'\right)\end{aligned}$$

Since at (x_0,y_0), $\nabla F = 0$, thus

$$P(x,y|d) \propto \exp\left(-\frac{1}{2}\begin{bmatrix} x-x_0 & y-y_0 \end{bmatrix} H|_{x_0,y_0}\begin{bmatrix} x-x_0 & y-y_0 \end{bmatrix}'\right)$$

write H as

$$H = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$$

thus the joint pdf will be

$$f(x,y) = K\exp\left(-\frac{1}{2}[A(x-x_0)^2 + 2B(x-x_0)(y-y_0) + C(y-y_0)^2]\right)$$

where K is the constant that normalize the pdf, to get it, we have

$$\begin{aligned}\int_{-\infty}^{\infty} dy\int_{-\infty}^{\infty} dx f(x,y) &= \int_{-\infty}^{\infty} dy\int_{-\infty}^{\infty} dx K\exp\left(-\frac{1}{2}[A(x-x_0)^2 + 2B(x-x_0)(y-y_0) + C(y-y_0)^2]\right)\\ &= \int_{-\infty}^{\infty} dy\int_{-\infty}^{\infty} dx K\exp\left(-\frac{1}{2}[Ax^2 + 2Bxy + Cy^2]\right)\\ &= \int_{-\infty}^{\infty} dy\int_{-\infty}^{\infty} dx K\exp\left(-\frac{1}{2}[A(x+By/A)^2 + (C-B^2/A)y^2]\right)\\ &= \int_{-\infty}^{\infty} dy K\exp\left(-\frac{1}{2}(C-B^2/A)y^2\right)\int_{-\infty}^{\infty} dx\exp\left(-\frac{1}{2}[A(x+By/A)^2]\right)\\ &= \int_{-\infty}^{\infty} dy K\exp\left(-\frac{1}{2}(C-B^2/A)y^2\right)\int_{-\infty}^{\infty} dx\exp\left(-\frac{1}{2}Ax^2\right)\\ &= \int_{-\infty}^{\infty} dy K\exp\left(-\frac{1}{2}(C-B^2/A)y^2\right)\sqrt{2\pi/A}\\ &= K\sqrt{2\pi/(C-B^2/A)}\sqrt{2\pi/A}\\ &= 2\pi K/\sqrt{AC-B^2}\\ &= 1\end{aligned}$$

Now we want to show $E[x] = x_0$. This is true because

$$\begin{aligned}
E[x - x_0] &= \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad (x - x_0) f(x,y) \\
&\propto \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad (x - x_0) \exp\left(-\frac{1}{2}[A(x-x_0)^2 + 2B(x-x_0)(y-y_0) + C(y-y_0)^2]\right) \\
&\propto \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad x \exp\left(-\frac{1}{2}[Ax^2 + 2Bx(y-y_0) + C(y-y_0)^2]\right) \\
&\propto \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad x \exp\left(-\frac{1}{2}[Ax^2 + 2Bxy + Cy^2]\right) \\
&\propto \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad x \exp\left(-\frac{1}{2}[A(x + By/A) + (C - B^2/A)y^2]\right) \\
&\propto \int_{-\infty}^{\infty} dy (By/A) \exp\left(-\frac{1}{2}[(C - B^2/A)y^2]\right) \\
&\propto \int_{-\infty}^{\infty} dy \quad y \exp\left(-\frac{1}{2}[(C - B^2/A)y^2]\right) \\
&= 0
\end{aligned}$$

Similarly $E[y] = y_0$.
Thus

$$\begin{aligned}
\sigma_x^2 &= E[(x - E[x])^2] \\
&= E[(x - x_0)^2] \\
&= \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad \frac{\partial f(x,y)}{\partial A} / (-1/2) \\
&= -2 \frac{\partial}{\partial A} \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad f(x,y) \\
&= -2 \frac{\partial}{\partial A} 2\pi K / \sqrt{AC - B^2} \\
&= \frac{2\pi K}{(A - B^2/C)\sqrt{AC - B^2}} \\
&= \frac{C}{AC - B^2}
\end{aligned}$$

Similarly

$$\sigma_y^2 = \frac{A}{AC - B^2}$$

$$\begin{aligned}
\sigma_{xy} &= E[(x - E[x])(y - E[y])] \\
&= E[(x - x_0)(y - y_0)] \\
&= \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad \frac{\partial f(x,y)}{\partial B} / (-1) \\
&= -\frac{\partial}{\partial B} \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \quad f(x,y) \\
&= \frac{-B}{AC - B^2}
\end{aligned}$$

In sum

$$\begin{aligned}
\boldsymbol{H}^{-1} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}^{-1} &= \frac{1}{AC - B^2} \begin{pmatrix} C & -B \\ -B & A \end{pmatrix} \\
&= \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix}
\end{aligned}$$

Problem 2 50 points

(b) (25 points)

Here are some topics people mentioned about the ABT book:

- Appendix A, background knowledge
- Number of data points vs number of model parameter, overdetermined, underdetermined, mixed determined
- MAP and Bayes' theorem
- L1 norm is better when existence of outlier
- difficulty in inverse: existance, uniqueness, and instability
- Fredholm integration equation of the first kind generalize many inverse problem
- Chapter 11 Section 3, general multivarate normal case with prior
- damped Newton's method, choose the step length instead of using full newton step
- ABT mentioned p-values, chi-square statistics
- in lecture, we maximize $P(d|m)$ to get the maximum likelihood, in ABT they defined $L(m|d) = P(d|m)$ first, then maximize it. ABT one is more clear, because $P(d|m)$ apprears to be a function of d, where in fact we want it to be a function of m.

(c) (25 points)

Here are some topics people mentioned:

- When estimating σ^2, devided by $N - 1$ instead of N
- Confused about $J(m)$ and $G(m)$, it's really the same thing, Jacobian
- Class did a better job of describing prior
- Class and ABT does not cover the case when prior is a given range
- numerical calculation of derivative when it's hard to do it analytically
- any method for finding global minimum?