INSTITUTO NACIONAL DE PESQUISA ESPACIAL – INPE

PROVA DISCURSIVA

TG36

DESENVOLVIMENTO OU APRIMORAMENTO DE SISTEMA DE ASSIMILAÇÃO DE DADOS NAS COMPONENTES DO SISTEMA TERRESTRE E DE APLICAÇÕES PARA MONITORAMENTO DO PROCESSO DE ASSIMILAÇÃO

SUA PROVA

- Além deste caderno contendo **5 (cinco)** questões discursivas **com as respectivas folhas de rascunho**, você receberá do fiscal de prova as folhas de textos definitivos;

TEMPO

- Você dispõe de **4 (quatro) horas** para a realização da prova;
- **2 (duas) horas** após o início da prova, é possível retirar-se da sala, sem levar o caderno de questões;
- A partir dos **30 (trinta) minutos** anteriores ao término da prova é possível retirar-se da sala **levando o caderno de questões**.

NÃO SERÁ PERMITIDO

- Qualquer tipo de comunicação entre os candidatos durante a aplicação da prova;
- Anotar informações relativas às respostas em qualquer outro meio que não seja no caderno de questões e nas folhas de textos definitivos;
- Levantar da cadeira sem autorização do fiscal de sala;
- Usar o sanitário ao término da prova, após deixar a sala.

INFORMAÇÕES GERAIS

- Verifique se seu caderno de questões está completo, sem repetição de questões ou falhas. Caso contrário, **notifique imediatamente o fiscal da sala**, para que sejam tomadas as devidas providências;
- Confira seus dados pessoais, especialmente nome, número de inscrição e documento de identidade e leia atentamente as instruções para preencher as folhas de textos definitivos;
- Para o preenchimento das folhas de textos definitivos, use somente caneta esferográfica, fabricada em material transparente, com tinta preta ou azul;
- Assine seu nome apenas no(s) espaço(s) reservado(s) no cartão de respostas;
- Caso você tenha recebido caderno de cargo **diferente** do impresso em suas folhas de textos definitivos, o fiscal deve ser **obrigatoriamente** informado para o devido registro na ata da sala;
- O preenchimento das folhas de textos definitivos é de sua responsabilidade e **não será permitida a troca de folha de texto definitivo em caso de erro cometido pelo candidato**;
- Para fins de avaliação, serão levadas em consideração apenas os textos das folhas de textos definitivos;
- A FGV coletará as impressões digitais dos candidatos na lista de presença;
- Os candidatos serão submetidos ao sistema de detecção de metais quando do ingresso e da saída de sanitários durante a realização das provas.
- **Boa prova!**

FGV CONHECIMENTO

Questão 1

De forma abrangente, do ponto de vista dos problemas inversos generalizados nos quais a função custo engloba a covariância dos erros do modelo, do *background* e das observações, assim como os respectivos operadores e variáveis adjuntos, **explique as diferenças e similaridades entre os métodos de assimilação 3DVAR e 4DVAR, respondendo às seguintes perguntas:**

A) Quais as simplificações e/ou aproximações que são aplicadas na formulação do 3DVAR e 4DVAR para solução do problema de assimilação? Explique o porquê dessas simplificações, e conforme o caso, qual a abordagem utilizada.

B) Em termos de solução da equação de análise, indique como os métodos variacionais, sejam o 3DVAR ou 4DVAR, se diferenciam dos métodos baseados na Interpolação Ótima (IO). Explique as desvantagens do IO em relação aos métodos variacionais. Exemplifique no caso da assimilação de dados de radiâncias.

C) Dentro da janela de tempo da assimilação, em que passo de tempo é computado o operador das observações no 3DVAR e 4DVAR?

D) De forma geral, quais as vantagens e desvantagens do 4DVAR em relação ao 3DVAR?

Questão 2

As técnicas de assimilação são amplamente utilizadas em áreas tais como meteorologia na redução de erro dos modelos numéricos. Existem várias técnicas para assimilação, sendo que uma das principais técnicas utiliza o filtro de Kalman.

Sobre o filtro de Kalman, responda ao que se pede a seguir.

A) O filtro de Kalman apresenta duas fases distintas. identifique-as e explique-as sucintamente.

B) O filtro de Kalman é um procedimento iterativo. No entanto, existem problemas práticos que podem levar o algoritmo a divergir. Enumere e explique dois desses problemas.

Questão 3

O método dos gradientes conjugados (MGC) é uma técnica iterativa para resolver sistemas lineares de grande porte, que é frequentemente empregada na assimilação de dados em métodos como o 3DVAR e 4DVAR, por exemplo. Um ponto de partida para derivar o MGC é através da função quadrática de teste:

$$\phi(x) = \frac{1}{2}x^T Ax - x^T b$$

em que $b, x \in R^p$, e a matriz $A \in R^{pxp}$é simétrica, positiva e definida. Para minimizarmos essa função, calculamos o gradiente da função no pontox^μonde o gradiente é zero:

$$grad\left(\phi(x_\mu)\right) = Ax_\mu - b = 0$$

ou seja,

$$Ax_\mu = b$$

Isso mostra que o mesmo$x_\mu$que minimiza a função $\phi(x)$também serve como solução de um sistema linear$Ax = b$, que nos traz motivação a criarmos um método capaz de minimizar ϕ ao invés de utilizarmos pacotes de álgebra linear.

Diante deste contexto, dado o algoritmo do MGC:

Começar com algum x_0. Iniciar $p_0 = r_0 = b - Ax_0$

Para k=0, 1, 2... até atingir o critério de parada

$\alpha_k = \frac{r_k^\top r_k}{p_k^\top A\ p_k}$

$x_{k+1} = x_k + \alpha_k p_k$

$r_{k+1} = r_k - \alpha_k A p_k$

$\beta_{k+1} = \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k}$

$p_{k+1} = r_{k+1} + \beta_{k+1} p_k$

Resultado final x_{k+1}

A) Considere as matrizes$A$ e $b$abaixo:

$$\text{A}=\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \qquad \text{b}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Para testar o algoritmo MGC, calcule explicitamente cada variável do algoritmo nos passos $k = 0$e $k = 1$, iniciando com

$$\text{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

B) Qual é o critério de parada do algoritmo para que as iterações terminem garantindo a convergência da solução na precisão desejada?

C) Indique maneiras desse algoritmo ser otimizado considerando que será realizado serialmente. Desconsidere o uso de pré condicionadores ou bibliotecas de processamento paralelo.

Questão 4

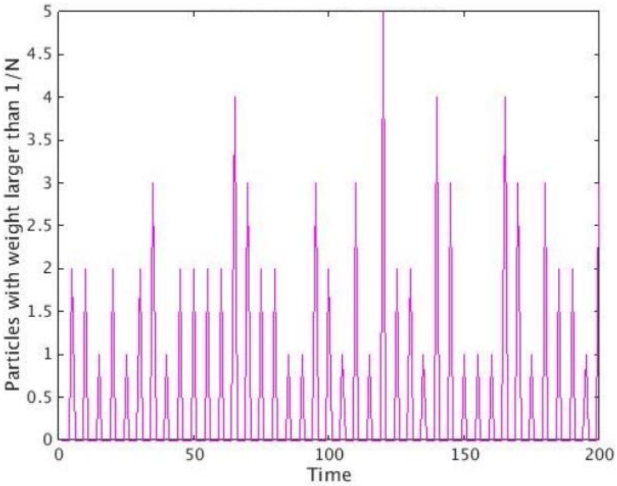
Um pesquisador gostaria de estudar o comportamento de um Filtro de Partículas e decidiu começar por um modelo simples 1-D (unidimensional) definido pela equação:

$$x^{n+1} = x^n + 20 \frac{x^n * x^n}{1 + \exp(x^n - 7)} + \beta^n$$

em que x é a variável a ser prevista, n é o índice de tempo e β^n é definido por um número randômico gerado a partir de um espaço amostral M(0,0.3). Após realizar uma rodada do modelo com 200 passos de tempo representando a realidade, o pesquisador gerou um programa no Matlab de amostragem estocástica universal (*"stochastic universal sampling"*), como um algoritmo padrão de Filtro de Partículas de reamostragem.

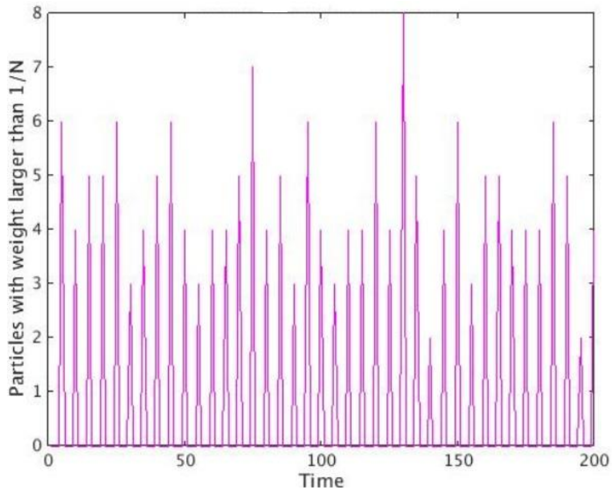
O código possuía observações disponíveis a cada 5 passos de tempo e contemplava um número N = 20 de partículas. Ao gerar uma plotagem que apresentava a quantidade de partículas com pesos acima de 1/N no momento da reamostragem, verificou-se o que a figura 1 abaixo mostra:

Figura 1



O pesquisador decidiu continuar seu estudo explorando a ideia de densidade de proposta ("proposal density"). Para tanto, alterou a equação de seu modelo, adicionando o seguinte termo à equação acima: $K(t)\ (y^n - x^n)$, onde K(t) varia linearmente de 0 a 4 entre as observações (y). Ao utilizar este artifício, gerou uma nova plotagem referente à quantidade de partículas, conforme mostra a figura 2 abaixo:

Figura 2



A) Explique como funciona um Filtro de Partículas simples baseado em Amostragem por Importância e de que maneira métodos de reamostragem tais como o "*stochastic universal sampling*" funcionam no filtro.

B) Analise as figuras 1 e 2, respondendo aos itens a seguir:

B_1 Qual das figuras se mostra mais promissora a representar melhores resultados e por que? Qual o problema principal encontrado na figura menos promissora?

B_2 Cite pelo menos 1 esquema de Filtro de Partículas da atualidade que venha apresentando resultados promissores em sistemas de grande dimensão não-lineares.

Questão 5

Considere que há N observações independentes, realizadas com instrumentos diferentes, de uma variável S em um ponto do espaço, denominadas S_n , $n = 1, N$. O erro de cada observação é dado por $\varepsilon_n = S_n - S$. A probabilidade de o erro estar entre ε_n e $\varepsilon_n + d\varepsilon_n$ é

$$p(\varepsilon_n) = \frac{1}{\sigma_{n\sqrt{2\pi}}} \exp\left[-\frac{\varepsilon_n^2}{2\sigma_n^2}\right]$$

em que $\sigma_n^2 = < \varepsilon_n^2 > = \int_{-\infty}^{\infty} \varepsilon_n^2\, p(\varepsilon_n) d\varepsilon_n$ é a variância conhecida de S_n.

Com base nesses dados responda ao que se pede a seguir.

A) Se $N = 1$, qual é o valor mais provável de S? Justifique sua resposta.

B) Se houver agora duas fontes de informação independentes sobre S, sendo uma, S_o, proveniente de observação, com $\varepsilon_o = S_o - S$, $\sigma_o^2 = < \varepsilon_o^2 >$, e outra, de um modelo S_b com $\varepsilon_b = S_b - S$, $\sigma_b^2 = < \varepsilon_b^2 >$, qual é o valor mais provável de S, isto é, qual é o valor da análise S_a? Discuta sua solução considerando os casos em que $\sigma_o^2 \gg \sigma_b^2$ e $\sigma_o^2 \ll \sigma_b^2$.

C) Para o caso em (B), qual é a estimativa da esperança do erro quadrático da análise $< \varepsilon_a^2 > = < (S_a - S)^2 >$?

D) Considere agora um conjunto de N estimativas independentes de S, mas que todas têm exatamente o mesmo σ^2 . Qual é a estimativa do $< \varepsilon_a^2 >$ e qual é o comportamento do erro da análise em função do aumento de estimativas disponíveis?

Realização

FGV CONHECIMENTO