



INSTITUTO NACIONAL DE PESQUISA ESPACIAL – INPE

PROVA OBJETIVA

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 **45 (quarenta e cinco)** questões objetivas, você receberá do fiscal de prova o cartão de respostas;
- As questões objetivas têm **5 (cinco)** opções de resposta (A, B, C, D e E) e somente uma delas está correta.



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 e também confira seu cargo. Caso tenha recebido caderno de cargo **diferente** do impresso em seu cartão de respostas, o fiscal deve ser **obrigatoriamente** informado para o devido registro na ata da sala;
- Confira seus dados pessoais, especialmente nome, número de inscrição e documento de identidade e leia atentamente as instruções para preencher o cartão de respostas;
- Para o preenchimento do cartão de respostas, 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;
- Reserve tempo suficiente para o preenchimento do seu cartão de respostas. O preenchimento é de sua responsabilidade e **não será permitida a troca do cartão de respostas em caso de erro cometido pelo candidato**;
- Para fins de avaliação, serão levadas em consideração apenas as marcações realizadas no cartão de respostas;
- 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!

CONHECIMENTOS ESPECÍFICOS

1

Podemos dizer que a Assimilação de Dados é um conjunto de técnicas empregadas para realizar adequadamente a inserção de dados de observação num sistema operacional de previsão, cujo propósito é

- (A) usar todas as informações disponíveis para determinar, com a maior precisão possível, o estado do fluxo meteorológico.
- (B) usar todas as informações disponíveis para determinar, com a maior precisão possível, o estado do fluxo atmosférico ou oceânico.
- (C) usar as informações amostrais para determinar, com a maior confiabilidade possível, o estado do fluxo meteorológico.
- (D) usar as informações amostrais para determinar, com a maior confiabilidade possível, o estado do fluxo atmosférico ou oceânico.
- (E) usar todas as informações experimentais para determinar, com a maior precisão possível, o estado do fluxo atmosférico ou oceânico.

2

No estudo da altura da superfície do mar, um especialista optou por utilizar um método estatístico multivariado como método de estimação.

Assinale a opção que indica o método que seria adequado.

- (A) Relaxamento Newtoniano.
- (B) Método de Correção Sucessiva.
- (C) Aproximação Variacional.
- (D) Assimilação Sequencial Simples.
- (E) Interpolação Ótima.

3

Um centro de estudos atmosféricos deseja descrever sistemas que surgem de aplicações ambientais.

Dessa forma, resolvem utilizar esquemas de assimilação de dados cujas propriedades mais significativas são modelos

- (A) tradicionais de estimação e lineares.
- (B) tradicionais de estimação e não lineares.
- (C) muito grandes e lineares.
- (D) muito grandes e não lineares.
- (E) simplificados e lineares.

4

Um pesquisador usou em seu trabalho para estimar a tendência central de certo fenômeno meteorológico o seguinte estimador:

$$\hat{\theta} = \frac{\sum_{i=1}^{n/2} X_i}{n}$$

Sabe-se que o fenômeno em questão segue uma distribuição Normal com média θ e desvio-padrão σ .

O viés do estimador é

- (A) -2θ .
- (B) 2θ .
- (C) $-\theta/2$.
- (D) $\theta/2$.
- (E) 0.

5

Os métodos de estimação estatísticos são muito utilizados na estimação de parâmetros de modelos.

Assim, dentro das propriedades dos bons estimadores, as mais desejáveis são

- (A) não tendenciosidade e variância mínima.
- (B) não tendenciosidade e consistência.
- (C) consistência e variância mínima.
- (D) consistência e suficiência.
- (E) variância mínima e suficiência.

6

No que diz respeito aos problemas de assimilação de dados para sistemas dinâmicos não lineares, assinale a opção que indica o esquema que dá a melhor estimativa linear da solução para o problema de assimilação de mínimos quadrados.

- (A) Filtro de Kalman.
- (B) Assimilação sequencial ótimo.
- (C) Filtros de conjunto.
- (D) Métodos de amostragem.
- (E) Assimilação variacional.

7

Um grupo de trabalho da área de sensoriamento remoto estuda modelos de previsão de eventos climáticos. Dessa forma, decidiram utilizar como modelo inicial o seguinte modelo:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

onde α e β são respectivamente os coeficientes linear e angular do modelo e ε_i os erros aleatórios.

No entanto, verificaram que a escala usada nas variáveis dependente e independente do modelo não estavam adequadas. Dessa forma, propuseram a seguinte mudança de escala: $Y_i^* = 10Y_i$ e $X_i^* = 5X_i$.

Sabendo que a estimativa do coeficiente angular obtido do modelo anterior foi igual a 2 e considerando o novo modelo gerado, o estimador de Mínimos Quadrados Ordinários de β^* é

- (A) 1.
- (B) 2.
- (C) 5.
- (D) 10.
- (E) 20.

8

No estudo de dispositivos semicondutores usados na construção de satélites, optou-se por utilizar o método da máxima verossimilhança na estimação dos parâmetros do estudo. Segundo os especialistas, esse método foi escolhido por apresentar boas propriedades.

Assinale a opção que apresenta as propriedades que pertencem ao método escolhido.

- (A) Não tendenciosidade e consistência.
- (B) Invariância e tem distribuição assintoticamente normal.
- (C) Não tendenciosidade e tem distribuição assintoticamente normal.
- (D) Invariância e tem distribuição normal independentemente do tamanho da amostra.
- (E) Assintoticamente não tendenciosos e tem distribuição normal independentemente do tamanho da amostra.

9

Uma empresa faz pesquisas na área ambiental. Sabe-se que o tempo entre secas (em anos) em determinada região no Brasil segue uma distribuição exponencial com parâmetro β .

Considere uma amostra de tamanho 5 cujos elementos são 15, 18, 20, 22 e 25.

Aplicando o método da máxima verossimilhança, o valor da estimativa de β é

- (A) 0,01.
- (B) 0,03.
- (C) 0,05.
- (D) 0,10.
- (E) 0,12.

10

Cientistas interessados em estimar os parâmetros de modelos de assimilação oceânica, utilizaram um método de estimação cujo objetivo é minimizar a soma dos quadrados das diferenças entre o valor estimado e valor real.

Diante do exposto, assinale a opção que apresenta o método que se enquadra na descrição do objetivo acima.

- (A) da máxima verossimilhança.
- (B) bayesiano.
- (C) do problema inverso.
- (D) dos momentos.
- (E) dos mínimos quadrados.

11

Uma pesquisa sobre a dispersão espacial do risco de ocorrência de um determinado fenômeno utilizou a estimação Bayesiana como método de estimação.

Sobre esse método de estimação, assinale a opção correta.

- (A) Considera inicialmente que o parâmetro a estimar é desconhecido.
- (B) Considera inicialmente que o parâmetro a estimar é conhecido levando em conta os dados coletados da amostra.
- (C) Considera inicialmente que o parâmetro é uma quantidade cuja variação pode ser descrita por uma distribuição de probabilidade somente depois da retirada da amostra.
- (D) Considera inicialmente que o parâmetro é uma quantidade cuja variação pode ser descrita por uma distribuição de probabilidade antes da retirada da amostra.
- (E) Considera inicialmente uma amostra aleatória que é retirada e uma população indexada pelo parâmetro a estimar, baseado nos valores observados da amostra.

12

Um instituto deseja estudar a incidência de certo evento em determinados intervalos de tempo. Seja X_1, X_2, \dots, X_n uma amostra aleatória independente e identicamente distribuída conforme uma distribuição de Poisson com parâmetro λ .

Sabe-se que λ tem distribuição Gama com parâmetros α e β e que $Y = \sum_i X_i$.

Então, a distribuição *a posteriori* de λ é

- (A) Poisson com parâmetro λ .
- (B) Poisson com parâmetro $n\lambda$.
- (C) Gama com parâmetros $(y + \alpha; \frac{\beta}{n\beta+1})$.
- (D) Gama com parâmetros $(y - \alpha; \frac{\beta}{n\beta-1})$.
- (E) Gama com parâmetros $(\alpha + \beta; \frac{(n\beta+1) \cdot \lambda}{\beta})$.

13

Algoritmos de estimação aplicados a assimilação de dados requerem a solução de um problema de otimização.

Assinale a opção que indica o método que pode ser considerado híbrido.

- (A) Regressão Quantílica.
- (B) Filtro de Kalman.
- (C) 3D-VAR.
- (D) BLUE.
- (E) Evolução Diferencial.

14

Pesquisadores da área de sistema de assimilação de dados nas componentes do sistema terrestre resolveram utilizar um método de minimização variacional utilizando o algoritmo 3D-VAR para encontrar a solução de um problema de otimização.

Sobre as propriedades numéricas do método utilizado, assinale a afirmativa correta.

- (A) Persistência, eficiência e confiabilidade.
- (B) Persistência, robustez e confiabilidade.
- (C) Eficiência, robustez e independência.
- (D) Eficiência, robustez e confiabilidade.
- (E) Eficiência, persistência e independência.

15

Foi ajustado um modelo para prever a precipitação média mensal de chuvas para determinada região em função da temperatura média mensal (\bar{T}) e da umidade média mensal (\bar{U}).

Suponha que o seguinte modelo foi obtido:

$$\bar{P} = 30,5 + \bar{T} + 3\bar{U}$$

Supondo que o desvio padrão da temperatura média mensal é igual a 5 e o desvio padrão da umidade média mensal é igual a 4.

A propagação dos erros da precipitação média mensal é

- (A) 9.
- (B) 11.
- (C) 13.
- (D) 15.
- (E) 17.

16

Os Filtros Bayesianos são assim chamados por basearem-se na aplicação do Teorema de Bayes, que relaciona distribuições de probabilidade a *priori* com distribuições de probabilidade a *posteriori*.

Há dois passos fundamentais para a estimação de estados, onde o primeiro passo está associado ao modelo dinâmico do sistema ou processo, enquanto o segundo passo está associado ao modelo de observações ou sensoriamento.

Neste contexto, os passos são denominados, respectivamente,

- (A) Treinamento e Classificação.
- (B) Predição e Correção.
- (C) Atenuação e Geração.
- (D) Propagação Inversa e Propagação Direta.
- (E) Marginalização e Condicionamento.

Atenção: o enunciado a seguir refere-se às quatro próximas questões.

Seja um modelo dinâmico discreto unidimensional de caminhada aleatória dado por:

$$x_k = x_{k-1} + q_{k-1}, \quad q_{k-1} \sim \mathcal{N}(0, Q)$$

$$y_k = x_k + r_k, \quad r_k \sim \mathcal{N}(0, R)$$

Em que x_k e y_k são, respectivamente, o estado a ser estimado e a medição no tempo k . As variáveis aleatórias q_k e r_k possuem distribuição normal com média nula e variâncias Q e R , respectivamente, ambas iguais a 1. Assuma, ainda, que a distribuição de probabilidade do estado no tempo k independe da distribuição de probabilidade dos estados anteriores (i.e., o sistema atende à propriedade de Markov).

Em um determinado instante de tempo $k - 1$, o estado estimado por um filtro de Kalman é dado por 2,5 e sua variância é estimada em 1,0.

No instante de tempo k , obtém-se uma medição igual a 3,1.

17

Nessas condições, antes de se agregar a informação proveniente da medição no instante de tempo k , a predição do estado para esse mesmo instante k será

- (A) 1,5.
- (B) 2,0.
- (C) 2,5.
- (D) 3,0.
- (E) 3,5.

18

Antes de agregar a informação proveniente da medição no tempo k , a predição da variância do estado, para esse mesmo instante k , será

- (A) 1,0.
- (B) 1,1.
- (C) 1,4.
- (D) 2,0.
- (E) 3,0.

19

Após agregar a informação proveniente da medição no tempo k , o valor estimado do estado para esse mesmo instante k será

- (A) 2,10.
- (B) 2,17.
- (C) 2,83.
- (D) 2,90.
- (E) 3,10.

20

Após se agregar a informação proveniente da medição no tempo k , o valor estimado da variância do estado para esse mesmo instante k será

- (A) $1/2$.
- (B) $2/3$.
- (C) 1.
- (D) $4/3$.
- (E) $3/2$.

21

Filtros Bayesianos são métodos usados para estimar o estado de um sistema dinâmico que seja observado por meio de medidas com incertezas. Entre os algoritmos utilizados para implementação de filtros Bayesianos, pode-se citar o Filtro de Kalman clássico, aplicável a sistemas de modelos lineares e com distribuições Gaussianas de probabilidade.

Nesse contexto, assinale a opção que indica uma das características do Filtro de Kalman clássico.

- (A) Usar expansões de Taylor para determinar modelos e aproximações Gaussianas para distribuições de probabilidade gerais.
- (B) Representar distribuições de probabilidade como um conjunto de amostras discretas de Monte Carlo com seus respectivos pesos.
- (C) Filtrar frequências indesejadas de sinais físicos contínuos no tempo, eliminando faixas determinadas pelo projetista que o implementa por meio dos ganhos de Kalman.
- (D) Usar grande poder computacional para calcular distribuições de probabilidades numericamente, evitando cálculos com equações analiticamente intratáveis.
- (E) Ser uma solução fechada para o problema de filtragem ou redução de incertezas, que não depende de aproximações numéricas por conta da menor complexidade computacional dos cálculos recursivos de médias e covariâncias Gaussianas.

22

A utilização de Filtros de Kalman clássicos (*Kalman Filters* - KF) ou estendidos (*Extended Kalman Filters* - EKF) para a assimilação de dados envolve dificuldades práticas.

Com relação a essas dificuldades, analise as afirmativas a seguir.

- I. O EKF é o método otimizado para a assimilação de dados sequencial de um modelo dinâmico linear n -dimensional, sendo o KF apropriado apenas para sistemas unidimensionais.
- II. O uso do KF e do EKF em modelos dinâmicos que contam com vetores de estados com muitas dimensões requer alta capacidade computacional e de armazenamento, tornando-os práticos apenas para modelos simplificados, de baixa dimensionalidade.
- III. A linearização de modelos não lineares envolve a aproximação de funções matemáticas com o truncamento de séries, o que pode gerar erros de propagação de covariâncias, especialmente em modelos de alta dimensionalidade.

Está correto o que se afirma em

- (A) I, apenas.
- (B) I e II, apenas.
- (C) I e III, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

23

O Filtro de Kalman por Conjunto, ou *Ensemble Kalman Filter* - EnKF, representa uma alternativa ao Filtro de Kalman Clássico (KF) e ao Filtro de Kalman Estendido (EKF) para a assimilação de dados sequencial com grandes conjuntos de dados.

Entre as vantagens do EnKF com relação ao KF e ao EKF, destaca-se a

- (A) redução da dimensionalidade dos estados do modelo, que permite a queda abrupta dos esforços computacionais para a assimilação recursiva.
- (B) aplicação de um método de Monte Carlo, que garante maior facilidade ao cálculo recursivo de propagação de covariâncias, que são aproximadas pela covariância de um conjunto de possíveis estados do modelo.
- (C) derivação e aplicação eficiente de um operador tangente linear, equivalente ao Jacobiano da função matemática associada ao modelo dinâmico do sistema.
- (D) formulação do método para cálculo recursivo de propagação de distribuições de probabilidades não-gaussianas.
- (E) aplicabilidade do método à estimação de estados de sistemas dinâmicos lineares e não-lineares.

24

Seja um modelo não linear dado por:

$$\begin{aligned} x_k &= M(x_{k-1}) + q_{k-1}, & q_{k-1} &\sim \mathcal{N}(0, Q); \\ y_k &= H(x_k) + r_k, & r_k &\sim \mathcal{N}(0, R), \end{aligned}$$

em que: x_k é um vetor de estados de n dimensões em um dado instante de tempo k ; M e H são mapeamentos não-lineares de R^n para R^n e de R^n para R^m , respectivamente; q e r são vetores aleatórios gaussianos de média nula e covariância Q e R , respectivamente.

Considere a implementação de um Filtro de Kalman por Conjunto (*Ensemble Kalman Filter* - EnKF) com 1000 pontos representando possíveis estados. Cada um dos 1000 pontos é denotado $x_t^{(i)}$, onde i é inteiro e varia de 1 a 1000.

Considere, ainda, que a média dos pontos do conjunto no instante k pode ser representada por $\bar{x}_k = \sum_{i=1}^{1000} \frac{x_k^{(i)}}{1000}$, e que o ganho de Kalman no instante k é geralmente representado pelo produto de uma matriz A pela inversa de uma matriz B ($K_k = AB^{-1}$).

Considerando as condições enunciadas acima, para garantir estimativas de covariâncias não enviesadas, a matriz A pode ser calculada pela expressão:

- (A) $\frac{1}{999} \sum_{i=1}^{1000} [x_k^{(i)} - \bar{x}_k] [H(x_k^{(i)}) - \bar{y}_k]^T$.
- (B) $\frac{1}{999} \sum_{i=1}^{999} [x_k^{(i)} - M(\bar{x}_k)] [H(x_k^{(i)}) - \bar{y}_k]^T$.
- (C) $\frac{1}{999} \sum_{i=1}^{999} [M(x_k^{(i)}) - \bar{x}_k] [H(x_k^{(i)}) - M(\bar{y}_k)]^T$.
- (D) $\frac{1}{999} \sum_{i=1}^{999} [x_k^{(i)} - \bar{x}_k] [H(x_k^{(i)}) - \bar{y}_k]^T$.
- (E) $\frac{1}{999} \sum_{i=1}^{1000} [x_k^{(i)} - M(\bar{x}_k)] [H(x_k^{(i)}) - \bar{y}_k]^T$.

25

Considere o modelo não linear e o Filtro de Kalman por Conjunto (EnKF) detalhados na questão 04.

Para garantir estimativas de covariâncias não enviesadas, a matriz B pode ser calculada pela expressão:

- (A) $\frac{1}{999} \sum_{i=1}^{1000} [M(x_k^{(i)}) - \bar{x}_k] [M(x_k^{(i)}) - \bar{x}_k]^T + Q$.
- (B) $\frac{1}{999} \sum_{i=1}^{999} [H(x_k^{(i)}) - \bar{y}_k] [H(x_k^{(i)}) - \bar{y}_k]^T + H(Q)$.
- (C) $\frac{1}{999} \sum_{i=1}^{1000} [H(x_k^{(i)}) - \bar{y}_k] [H(x_k^{(i)}) - \bar{y}_k]^T + R$.
- (D) $\frac{1}{999} \sum_{i=1}^{999} [M(x_k^{(i)}) - \bar{x}_k] [M(x_k^{(i)}) - \bar{x}_k]^T + R$.
- (E) $\frac{1}{999} \sum_{i=1}^{999} [H(x_k^{(i)}) - M(\bar{x}_k)] [H(x_k^{(i)}) - M(\bar{x}_k)]^T$.

26

Filtros de Partículas são implementações não paramétricas de filtros Bayesianos em que as distribuições de probabilidade não são explicitamente definidas, sendo, portanto, representadas por um conjunto de amostras provenientes delas próprias (denominadas partículas).

Com relação aos filtros de partículas, analise as afirmativas a seguir e assinale (V) para a verdadeira e (F) para a falsa.

- () As partículas representam observações (ou medidas) obtidas por sensores aplicados ao sistema em análise, e a elas são associados pesos proporcionais às suas probabilidades de coincidirem com medidas correspondentes ao estado verdadeiro do sistema.
- () Quando aplicados à assimilação de dados, a cada passo de assimilação, novos pesos são atribuídos às partículas. Caso não seja realizado nenhum processo de reamostragem, o conjunto de partículas costuma degenerar-se, com uma das partículas recebendo peso normalizado próximo de 1 e as outras partículas recebendo pesos normalizados próximos de 0.
- () São capazes de representar distribuições de probabilidade multimodais, isto é, cujas densidades de probabilidade possuem mais de um máximo local.

As afirmativas são, respectivamente,

- (A) F – F – V.
- (B) V – V – F.
- (C) V – V – V.
- (D) F – V – V.
- (E) F – V – F.

27

Filtros de partículas são, em geral, implementados com o uso de reamostragem sequencial por importância. Essa reamostragem pode ser adaptativa, ocorrendo apenas quando a métrica denominada número efetivo de partículas é considerada muito baixa.

Considerando um filtro de partículas com N partículas cujos pesos são dados por $w_{(i)}$, $i = 1, \dots, N$, a estimativa do número efetivo de partículas é dada por

- (A) $\frac{1}{\sum_{i=1}^N w_{(i)}}$.
- (B) $\frac{1}{\sum_{i=1}^N w_{(i)}^2}$.
- (C) $\sum_{i=1}^N \frac{1}{w_{(i)}}$.
- (D) $\sum_{i=1}^N \frac{1}{w_{(i)}}$.
- (E) $\frac{\sum_{i=1}^N w_{(i)}^2}{\sum_{i=1}^N w_{(i)}^3}$.

28

A reamostragem em filtros de partículas pode ser realizada por meio da criação de novas amostras retiradas das distribuições de probabilidade discretas correspondentes a conjuntos de partículas e suas configurações de pesos. No entanto, o fato de as novas amostras serem criadas exatamente nos mesmos pontos do espaço em que se localizam as partículas anteriores é inconveniente, pois facilita o empobrecimento das partículas (i.e., o chamado *particle impoverishment*).

Uma forma de produzir um novo conjunto de partículas em pontos distintos é substituir as distribuições discretas de probabilidade por aproximações contínuas e, somente então, realizar a reamostragem. A criação dessas aproximações se dá por meio de uma operação matemática entre a distribuição de probabilidade discreta e um *kernel* contínuo.

Nesse contexto, o processo de reamostragem em distribuições de probabilidade contínuas, que aproximam distribuições discretas correspondentes às configurações de partículas, é chamado de

- (A) Convolução.
- (B) Suavização.
- (C) Integração.
- (D) Nuclearização.
- (E) Regularização.

29

O problema de previsão numérica de tempo em escala global é de altíssima dimensionalidade, envolvendo, por exemplo, representações de estados com centenas de milhões de variáveis.

Essa alta dimensionalidade impõe grandes dificuldades para a aplicação de filtros de partículas (PF) em problemas de assimilação de dados com muitas observações independentes, porque nessas situações o número de partículas necessárias para representar as distribuições de probabilidade cresce exponencialmente.

Técnicas recentemente desenvolvidas que visam contornar essas dificuldades baseiam-se em combinar filtros de partículas e filtros de Kalman por conjunto (EnKF), criando-se soluções híbridas PF-EnKF.

Assinale a opção que indica a principal vantagem de se utilizar filtros híbridos PF-EnKF.

- (A) A redução do número de partículas necessárias para representar distribuições, que passam a ser caracterizadas apenas pelo primeiro e pelo segundo momentos.
- (B) A seleção de faixas específicas das variáveis de estado, reduzindo a dimensionalidade momentaneamente a cada passo de assimilação.
- (C) Sua adequação a distribuições não gaussianas quando necessário, ao mesmo tempo que se tenta manter a eficiência dos EnKF quando as distribuições são gaussianas.
- (D) Sua convergência mais rápida, em que uma partícula se sobressai com relação às demais em poucos passos de assimilação.
- (E) Sua maior simplicidade de implementação, pois o processo de reamostragem por importância torna-se uma etapa desnecessária nos passos de assimilação, podendo ser evitado.

30

Métodos de assimilação de dados clássicos são tradicionalmente classificados em sequenciais ou variacionais. Os métodos variacionais guardam semelhanças com a teoria de controle ótimo, por sua vez desenvolvida a partir do estabelecimento dos fundamentos do cálculo variacional.

Com relação à formulação variacional de assimilação de dados, analise as afirmativas a seguir e assinale (V) para a verdadeira e (F) para a falsa.

- () Trata-se da busca por estados dos sistemas que minimizam um funcional de custo, em geral definido como um erro quadrático entre observações e previsões correspondentes àqueles estados, calculadas por modelos matemáticos.
- () Envolve a necessidade de aplicação de técnicas de localização e/ou inflação de covariâncias para eliminar correlações espúrias entre possíveis soluções de problemas de otimização.
- () Baseia-se em otimizações com restrições dinâmicas fortes, introduzidas no problema por uso de multiplicadores de Lagrange; ou fracas, introduzidas no problema como termos ponderados de penalidades.

As afirmativas são, respectivamente,

- (A) F – V – V.
- (B) V – F – V.
- (C) V – F – F.
- (D) F – F – V.
- (E) F – V – F.

31

Relacione os algoritmos de otimização utilizados em assimilação de dados variacional com suas respectivas características correspondentes.

1. Método de Newton
2. Broyden-Fletcher-Goldfarb-Shanno (BFGS)
3. Gradiente Conjugado

- () Determina pontos cada vez mais próximos das soluções dos problemas de otimização mudando a direção de busca a cada iteração.
- () Requer o cálculo das expressões fechadas dos gradientes e matrizes Hessianas a cada iteração.
- () Utiliza aproximações de matrizes Hessianas e suas inversas para reduzir a carga computacional a cada iteração.

Assinale a opção que indica a relação correta, segundo a ordem apresentada.

- (A) 3 – 1 – 2.
- (B) 1 – 2 – 3.
- (C) 2 – 1 – 3.
- (D) 3 – 2 – 1.
- (E) 2 – 3 – 1.

32

Em assimilação variacional, frequentemente são encontrados problemas inversos mal-postos, (*ill-posed problems*). Esses problemas podem ser convertidos em bem-postos (*well-posed*) pelo uso de técnicas de regularização. Um exemplo é o uso da regularização de Tikhonov, em que se adiciona um termo de regularização a um funcional a ser minimizado, evitando-se assim instabilidades numéricas durante o cálculo da solução.

Por exemplo: suponha que se busque um vetor x que resolva o sistema $Hx = y$, minimizando-se o funcional

$$J = \|Hx - y\|_2^2,$$

em que $\|\cdot\|_2$ é a norma L^2 (isto é, um problema de mínimos quadrados mal-posto). Pode-se adicionar o termo de regularização de Tikhonov ao funcional, substituindo-o por

$$J_\alpha = \|Hx - y\|_2^2 + \|\Gamma x\|_2^2,$$

em que $\Gamma = \alpha I$, e I é a matriz identidade.

Considere um caso hipotético onde as variáveis H , y e α possuem os seguintes valores:

$$H = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}, y = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \alpha = 1.$$

Neste caso, o vetor x que minimiza J_α é:

(A) $\begin{bmatrix} 2 \\ 21 \\ 4 \\ 21 \end{bmatrix}$.

(B) $\begin{bmatrix} 1 \\ 21 \\ 2 \\ 21 \end{bmatrix}$.

(C) $\begin{bmatrix} 1 \\ 7 \\ 2 \\ 7 \end{bmatrix}$.

(D) $\begin{bmatrix} 4 \\ 14 \\ 8 \\ 14 \end{bmatrix}$.

(E) $\begin{bmatrix} 4 \\ 21 \\ 8 \\ 21 \end{bmatrix}$.

33

Modelos de previsão meteorológica utilizados em esquemas de assimilação variacional como o 4D-VAR são, em geral, não-lineares e complexos, implementados em grandes quantidades de linhas de código de software. A implementação dos esquemas pode requerer, ainda, os cálculos do Modelo Tangente Linear (*Tangent Linear Model* - TLM) e do Modelo Adjunto, os quais podem, frequentemente, ser computados de maneira eficiente.

O conjunto de técnicas utilizadas para computação eficiente dos modelos mencionados constituem ferramentas de

- (A) Análise de Componentes Principais (*Principal Component Analysis*).
- (B) Previsão Numérica (*Numerical Prediction*).
- (C) Propagação Direta (*Forward Propagation*).
- (D) Diferenciação Automática (*Automatic Differentiation*).
- (E) Estimativa Estatística (*Statistical Estimation*).

34

Recentemente, tem-se observado o aumento dos usos de algoritmos de Inteligência Artificial (IA) aplicados à assimilação de dados. Muitos algoritmos de IA em assimilação são baseados em redes neurais e redes neurais profundas, que necessitam de uma etapa de treinamento.

Essas etapas de treinamento nem sempre são de fácil execução. Por exemplo, há um fenômeno que ocorre quando um algoritmo é treinado e apresenta bom desempenho para um conjunto particular de dados usado para treinamento, mas falha ao prever respostas para dados de entrada não incluídos naquele conjunto.

A esse fenômeno dá-se o nome, em inglês, de

- (A) Supertraining.
- (B) Underfitting.
- (C) Overfitting.
- (D) Reinforcement Learning.
- (E) Kerneling.

35

Redes neurais artificiais são elementos fundamentais para o uso de técnicas de aprendizado de máquina. São constituídas por camadas de unidades de processamento, chamadas de neurônios.

Relacione os tipos de redes neurais listadas a seguir, às suas principais características.

1. Redes de Propagação Direta (feedforward).
 2. Redes Neurais Recorrentes.
 3. Redes de Funções de Base Radial.
 4. Redes Auto-Organizáveis de Kohonen.
- () Rede que possui realimentação, de forma que as saídas são direcionadas para as entradas, formando-se um loop.
 - () Rede em que os sinais fluem apenas em uma direção, da entrada para a saída, exceto quando em treinamento.
 - () Rede que é treinada com aprendizado não supervisionado, criando clusters dos dados de entrada.
 - () Rede usada para aproximar funções contínuas a partir de combinações lineares de Gaussianas.

Assinale a opção que indica a relação correta na ordem apresentada.

- (A) 2 – 1 – 3 – 4.
- (B) 3 – 1 – 4 – 2.
- (C) 1 – 2 – 3 – 4.
- (D) 2 – 1 – 4 – 3.
- (E) 1 – 2 – 4 – 3.

36

As técnicas de aprendizado de máquina aplicadas à assimilação podem ser utilizadas de diversas maneiras para tratamento de dados. Um exemplo de processo que pode ser vantajoso para os algoritmos de assimilação é o de redução da dimensionalidade de um conjunto de dados, no qual se aplica treinamento não supervisionado para gerar representações “compactadas” das entradas originais. Esse processo permite a assimilação de dados no espaço latente, melhorando a eficiência de treinamento dos algoritmos.

Determinadas arquiteturas de rede neural são utilizadas para redução de dimensionalidade e para a geração de representações de dados no espaço latente, em que se destaca a arquitetura do tipo

- (A) Autoencoders.
- (B) Redes Neurais Bayesianas.
- (C) Redes Neurais Convolucionais.
- (D) Redes Adversariais Generativas.
- (E) Máquinas de Boltzmann Restritas.

37

Assimilação de dados profunda (*Deep Data Assimilation - DDA*) é uma técnica recente que integra aprendizado profundo e assimilação.

Utiliza-se uma rede neural recorrente para aprender o processo de assimilação, que por sua vez é treinada a partir dos estados de um sistema dinâmico e de seus resultados de assimilação correspondentes. Tais redes neurais recorrentes são implementadas com o uso de funções de ativação, que introduzem não linearidades às saídas dos neurônios das redes.

Assinale a opção que menos se adequa às características esperadas para funções de ativação.

- (A) Softmax.
- (B) Sigmoid.
- (C) Transformação Afim.
- (D) Tangente Hiperbólico.
- (E) Unidade Linear Retificada (*Rectified Linear Unit – ReLU*).

38

Algoritmos para assimilação de dados podem ser implementados de maneira eficiente e otimizada por meio de paralelização de processos.

O *Parallel Data Assimilation Framework* (PDAF) é um pacote de *software* que simplifica a implementação de métodos de assimilação, provendo versões totalmente paralelizadas de algoritmos, como por exemplo, diferentes versões dos Filtros de Kalman por conjunto (EnKF). Um dos requisitos de funcionamento do PDAF é o uso de um protocolo padronizado de comunicação para computação paralela.

O principal padrão de comunicação entre os processos paralelos executados em um sistema de memória distribuída, é denominado

- (A) *Secure Shell* (SSH).
- (B) *Message Passing Interface* (MPI).
- (C) *Transmission Control Protocol* (TCP).
- (D) *Representational State Transfer* (REST).
- (E) *Lightweight Communications and Marshalling* (LCM).

39

A integração de técnicas de inteligência artificial e aprendizado de máquina com assimilação de dados pode aumentar a confiabilidade das previsões por introduzir modelos orientados a dados obtidos por meio de observações. A performance dos modelos de aprendizado de máquina pode ser medida por algumas métricas, como por exemplo a métrica *Mean Absolute Error* (MAE).

Considere um modelo de regressão usado para prever valores de uma variável, conforme a tabela a seguir.

Valores previstos pelo modelo	Valores reais correspondentes
5,0	5,2
5,5	5,2
4,5	5,0
4,0	3,8
4,8	4,5

O MAE para o conjunto de dados representado na tabela será

- (A) -0,1.
- (B) -0,02.
- (C) 0,02.
- (D) 0,102.
- (E) 0,3.

40

Uma abordagem recentemente desenvolvida para estimar concentrações de agentes poluentes na atmosfera busca integrar assimilação de dados variacional e processos Gaussianos para aperfeiçoar previsões. Processos Gaussianos são modelos estatísticos usados em aprendizado de máquina para descrever observações feitas em algum domínio contínuo, tal como espaço ou tempo.

Com relação aos processos Gaussianos para aprendizado de máquina, analise as afirmativas a seguir e assinale (V) para a verdadeira e (F) para a falsa.

- () São uma generalização de dimensão infinita das distribuições normais multivariáveis, em que cada variável aleatória está diretamente relacionada a algum ponto do domínio contínuo considerado.
- () São amplamente utilizados em modelos de regressão, em que se busca prever a forma de uma função contínua incorporando-se informações provenientes de observações.
- () Podem ser utilizados em tarefas de classificação, em que se busca prever a probabilidade de um conjunto de dados de entrada pertencer a uma classe específica.

As afirmativas são, respectivamente,

- (A) V – V – F.
- (B) V – F – F.
- (C) V – V – V.
- (D) F – F – F.
- (E) F – F – V.

41

Aplicações de assimilação de dados em oceanografia produzem análises e previsões de diferentes variáveis que caracterizam o estado dos oceanos. O Escritório de Meteorologia do Reino Unido, mais conhecido como *Met Office*, utiliza um modelo oceanográfico global, chamado *Forecasting Ocean Assimilation Model* (FOAM), na implementação de um esquema de assimilação variacional tridimensional, conhecido como NEMOVAR.

Entre as variáveis que fazem parte do vetor de estados utilizado no NEMOVAR, encontram-se

- (A) salinidade (S) e pressão atmosférica (P).
- (B) radiação solar (R) e altura da superfície do mar (η).
- (C) concentração de nitrogênio (C_N) e posição lunar relativa (x_L, y_L, z_L).
- (D) acidez/alcalinidade (pH) e concentração de biodiversidade marinha (C_b).
- (E) velocidades horizontais das correntes oceânicas (u, v) e temperatura (T).

42

A área de assimilação de dados acoplados tem recebido atenção crescente de pesquisadores e tecnólogos interessados em previsão numérica de tempo. Nos esquemas de assimilação acoplados, dois ou mais modelos geofísicos são combinados, frequentemente utilizando também técnicas diferentes de assimilação.

Com relação à área de assimilação de dados acoplados, analise as afirmativas a seguir.

- I. Busca-se utilizar os modelos geofísicos simultaneamente, de forma a produzirem previsões/análises consistentes entre si.
- II. Busca-se modelar e analisar as interações da atmosfera com os solos, com os oceanos e com as geleiras do planeta, melhorando assim as capacidades de previsão numérica.
- III. Combinam-se os modelos geofísicos, de forma que os resultados de previsão de um modelo sirvam de condição de contorno para a solução do(s) outro(s) modelo(s) a ele combinado(s).

Está correto o que se afirma em

- (A) I, apenas.
- (B) I e II, apenas.
- (C) I e III, apenas.
- (D) II e III, apenas.
- (E) I, II e III.

43

O treinamento de algoritmos de inteligência artificial no desenvolvimento de aplicações para assimilação de dados meteorológicos exige o uso de bases de dados representativas de estados atmosféricos. Embora bases de dados sintéticas sejam úteis para treinamento, o uso de bases de dados reais é sempre preferível.

Assinale a opção que apresenta a base de dados real que descreve propriedades físicas de uma grande quantidade de situações atmosféricas, e que é utilizada para o treinamento de modelos de temperatura atmosférica.

- (A) *Climate Atmospheric Trends Dataset* (CATD).
- (B) *Global Thermographic Profiles Logger* (GTPL).
- (C) *Thermodynamic Initial Guess Retrieval* (TIGR).
- (D) *Satellite Temperature Scans Dataset* (SaTScan).
- (E) *Thermometric Atmosphere Research Dataset* (ThermoARD).

44

Considere a lista de processos computacionais abaixo.

1. Integração do modelo não-linear para frente no tempo.
2. Integração do modelo não-linear para trás no tempo.
3. Integração do modelo adjunto para frente no tempo.
4. Integração do modelo adjunto para trás no tempo.
5. Integração do modelo tangente linear no loop interno.
6. Integração do modelo tangente linear no loop externo.

Assinale a opção que apresenta os processos realizados em assimilação de dados 4DVAR incremental, com restrição forte, na sequência correta de execução.

- (A) 3 – 2 – 6.
(B) 1 – 4 – 6.
(C) 2 – 3 – 5.
(D) 1 – 4 – 5.
(E) 3 – 2 – 5.

45

Algoritmos para assimilação de dados geralmente envolvem cálculos complexos que dependem de diversos fatores, como o tamanho dos espaços de estados, número de pontos da grade em questão, tamanho da janela de assimilação, etc. Frequentemente, observa-se que dois algoritmos usados para solucionar um mesmo problema podem ter eficiências diferentes, por conta de diferenças em suas implementações.

Uma maneira de se mensurar e representar a complexidade de um algoritmo é contabilizar o número de operações de ponto-flutuante (*flops*) necessárias para executá-lo e utilizar a notação “O-grande”.

Considere o algoritmo a seguir, implementado em uma linguagem de pseudocódigo autoexplicativa.

```
1: // Sendo M uma matriz quadrada n x n
2:
3: linhas, colunas = formato(M);
4:
5: n = linhas;
6:
7: soma = 0;
8:
9: para i = 1 até n:
10:     para j = 1 até n:
11:         soma = soma + M[i][j]*M[i][j];
```

A complexidade desse algoritmo será

- (A) $O(2n^2)$.
(B) $O(2n^2 - 1)$.
(C) $O(n^2)$.
(D) $O(n)$.
(E) $O(n^2 - n - 1)$.

Realização

