Seleção de Parâmetros de Regularização por Minimização de Distâncias Esperadas de Bregman

Elias S. Helou, Lucas E. A. Simões, Sandra A. Santos
ICMC - USP / IMECC - UNICAMP

19 de setembro 2023

Problemas bem postos

$$\boldsymbol{b} = A\boldsymbol{x}$$

Problemas bem postos

$$\boldsymbol{b} = A\boldsymbol{x}$$

Definição de Hadmard

Problemas bem postos

$$\boldsymbol{b} = A\boldsymbol{x}$$

Definição de Hadmard

- Solução existe para todo $\boldsymbol{b}$

Problemas bem postos

$$\boldsymbol{b} = A\boldsymbol{x}$$

Definição de Hadmard

- Solução existe para todo $\boldsymbol{b}$
- Solução sempre é única

Problemas bem postos

$$\boldsymbol{b} = A\boldsymbol{x}$$

Definição de Hadmard

- Solução existe para todo $\boldsymbol{b}$
- Solução sempre é única
- Solução depende continuamente dos dados

Problemas bem postos

$$\boldsymbol{b} = A\boldsymbol{x}$$

Definição de Hadmard

- Solução existe para todo $\boldsymbol{b}$
- Solução sempre é única
- Solução depende continuamente dos dados

$$\boldsymbol{x}^{\dagger} = A^{\dagger}\boldsymbol{b}$$

Bem posto $\times$ mal condicionado

Em dimensão finita

Bem posto × mal condicionado

Em dimensão finita

- Melhor aproximação $\boldsymbol{x}^{\dagger}$ existe para todo $\boldsymbol{b}$

Bem posto × mal condicionado

Em dimensão finita

- Melhor aproximação $\boldsymbol{x}^{\dagger}$ existe para todo $\boldsymbol{b}$
- Melhor aproximação $\boldsymbol{x}^{\dagger}$ sempre é única

Bem posto × mal condicionado

Em dimensão finita

- Melhor aproximação $\boldsymbol{x}^{\dagger}$ existe para todo $\boldsymbol{b}$
- Melhor aproximação $\boldsymbol{x}^{\dagger}$ sempre é única
- Melhor aproximação $\boldsymbol{x}^{\dagger}$ depende continuamente dos dados

Bem posto $\times$ mal condicionado

Em dimensão finita

- Melhor aproximação $\boldsymbol{x}^\dagger$ existe para todo $\boldsymbol{b}$
- Melhor aproximação $\boldsymbol{x}^\dagger$ sempre é única
- Melhor aproximação $\boldsymbol{x}^\dagger$ depende continuamente dos dados
- Operador A é aproximação de operador compacto em dimensão infinita

Bem posto × mal condicionado

Em dimensão finita

- Melhor aproximação $\boldsymbol{x}^\dagger$ existe para todo $\boldsymbol{b}$
- Melhor aproximação $\boldsymbol{x}^\dagger$ sempre é única
- Melhor aproximação $\boldsymbol{x}^\dagger$ depende continuamente dos dados
- Operador A é aproximação de operador compacto em dimensão infinita
- Operador A torna-se muito mal condicionado

Bem posto $\times$ mal condicionado

Em dimensão finita

- Melhor aproximação $\boldsymbol{x}^\dagger$ existe para todo $\boldsymbol{b}$
- Melhor aproximação $\boldsymbol{x}^\dagger$ sempre é única
- Melhor aproximação $\boldsymbol{x}^\dagger$ depende continuamente dos dados
- Operador A é aproximação de operador compacto em dimensão infinita
- Operador A torna-se muito mal condicionado

$$\boldsymbol{x}_\epsilon^\dagger = A^\dagger(\boldsymbol{b} + \epsilon)$$

Bem posto × mal condicionado

Em dimensão finita

- Melhor aproximação $\boldsymbol{x}^\dagger$ existe para todo $\boldsymbol{b}$
- Melhor aproximação $\boldsymbol{x}^\dagger$ sempre é única
- Melhor aproximação $\boldsymbol{x}^\dagger$ depende continuamente dos dados
- Operador A é aproximação de operador compacto em dimensão infinita
- Operador A torna-se muito mal condicionado

$$\boldsymbol{x}_\epsilon^\dagger = A^\dagger(\boldsymbol{b} + \epsilon)$$

$$\frac{\|\boldsymbol{x}_\epsilon^\dagger - \boldsymbol{x}^\dagger\|}{\|\boldsymbol{x}^\dagger\|} \approx \mathrm{cond}(A)\frac{\|\epsilon\|}{\|\boldsymbol{b}\|}$$

Melhorando o condicionamento

Tikhonov

Melhorando o condicionamento

Tikhonov

$$\boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|A\boldsymbol{x} - (\boldsymbol{b} + \epsilon)\|^2 + \gamma\|\boldsymbol{x}\|^2\}$$

Melhorando o condicionamento

Tikhonov

$$\boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|A\boldsymbol{x} - (\boldsymbol{b}+\epsilon)\|^2 + \gamma\|\boldsymbol{x}\|^2\}$$

- $\lim_{\gamma\to 0} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \boldsymbol{x}^{\dagger}_{\epsilon}$

Melhorando o condicionamento

Tikhonov

$$\boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|A\boldsymbol{x} - (\boldsymbol{b} + \epsilon)\|^2 + \gamma\|\boldsymbol{x}\|^2\}$$

- $\lim_{\gamma\to 0} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \boldsymbol{x}^{\dagger}_{\epsilon}$
- $\lim_{\gamma\to\infty} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = 0$

Melhorando o condicionamento

Tikhonov

$$\boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|A\boldsymbol{x} - (\boldsymbol{b}+\epsilon)\|^2 + \gamma\|\boldsymbol{x}\|^2\}$$

- $\lim_{\gamma\to 0} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \boldsymbol{x}^{\dagger}_{\epsilon}$
- $\lim_{\gamma\to \infty} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = 0$
- Se γ é muito pequeno, condicionamento ruim

Melhorando o condicionamento

Tikhonov

$$\boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|A\boldsymbol{x} - (\boldsymbol{b}+\epsilon)\|^2 + \gamma\|\boldsymbol{x}\|^2\}$$

- $\lim_{\gamma\to 0} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \boldsymbol{x}^{\dagger}_{\epsilon}$
- $\lim_{\gamma\to \infty} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = 0$
- Se γ é muito pequeno, condicionamento ruim
- Se γ é muito grande, precisão ruim

Melhorando o condicionamento

Tikhonov

$$\boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\{\|A\boldsymbol{x} - (\boldsymbol{b} + \epsilon)\|^2 + \gamma\|\boldsymbol{x}\|^2\}$$

- $\lim_{\gamma\to 0} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = \boldsymbol{x}^{\dagger}_{\epsilon}$
- $\lim_{\gamma\to \infty} \boldsymbol{x}^{\dagger}_{\epsilon,\gamma} = 0$
- Se γ é muito pequeno, condicionamento ruim
- Se γ é muito grande, precisão ruim
- Como escolher γ?

Melhorando o condicionamento

Tikhonov

$$\boldsymbol{x}_{\epsilon,\gamma}^{\dagger} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|A\boldsymbol{x} - (\boldsymbol{b} + \epsilon)\|^2 + \gamma\|\boldsymbol{x}\|^2\}$$

- $\lim_{\gamma\to 0} \boldsymbol{x}_{\epsilon,\gamma}^{\dagger} = \boldsymbol{x}_{\epsilon}^{\dagger}$
- $\lim_{\gamma\to \infty} \boldsymbol{x}_{\epsilon,\gamma}^{\dagger} = 0$
- Se γ é muito pequeno, condicionamento ruim
- Se γ é muito grande, precisão ruim
- Como escolher γ?

$$\gamma^* = \arg\min_{\gamma\in[0,\infty)} \|\boldsymbol{x}_{\epsilon,\gamma}^{\dagger} - \boldsymbol{x}\|^2$$

Melhorando o condicionamento

Regularização

Melhorando o condicionamento

Regularização

- $\{B_\gamma\}_{\gamma \in \mathbb{R}_+}$

Melhorando o condicionamento

Regularização

- $\{B_\gamma\}_{\gamma \in \mathbb{R}_+}$
- $\lim_{\gamma \to 0} B_\gamma = A^\dagger$

Melhorando o condicionamento

Regularização

- $\{B_\gamma\}_{\gamma \in \mathbb{R}_+}$
- $\lim_{\gamma \to 0} B_\gamma = A^\dagger$
- $\lim_{\gamma \to \infty} B_\gamma = 0$

Melhorando o condicionamento

Regularização

- $\{B_\gamma\}_{\gamma \in \mathbb{R}_+}$
- $\lim_{\gamma \to 0} B_\gamma = A^\dagger$
- $\lim_{\gamma \to \infty} B_\gamma = 0$
- Se γ é muito pequeno, condicionamento ruim

Melhorando o condicionamento

Regularização

- $\{B_\gamma\}_{\gamma \in \mathbb{R}_+}$
- $\lim_{\gamma \to 0} B_\gamma = A^\dagger$
- $\lim_{\gamma \to \infty} B_\gamma = 0$
- Se γ é muito pequeno, condicionamento ruim
- Se γ é muito grande, precisão ruim

Melhorando o condicionamento

Regularização

- $\{B_\gamma\}_{\gamma \in \mathbb{R}_+}$
- $\lim_{\gamma \to 0} B_\gamma = A^\dagger$
- $\lim_{\gamma \to \infty} B_\gamma = 0$
- Se γ é muito pequeno, condicionamento ruim
- Se γ é muito grande, precisão ruim
- Como escolher γ?

Melhorando o condicionamento

Regularização

- $\{B_\gamma\}_{\gamma \in \mathbb{R}_+}$
- $\lim_{\gamma \to 0} B_\gamma = A^\dagger$
- $\lim_{\gamma \to \infty} B_\gamma = 0$
- Se γ é muito pequeno, condicionamento ruim
- Se γ é muito grande, precisão ruim
- Como escolher γ?

$$\gamma^* = \arg \min_{\gamma \in [0,\infty)} \|B_\gamma(\boldsymbol{b} + \boldsymbol{\epsilon}) - \boldsymbol{x}\|^2$$

Risco esperado

Modelo probabilístico

Risco esperado

Modelo probabilístico

- ϵ muitas vezes é um vetor de variáveis aleatórias

Risco esperado

Modelo probabilístico

- ϵ muitas vezes é um vetor de variáveis aleatórias
- O modelo de probabilidades é conhecido com precisão razoável

Risco esperado

Modelo probabilístico

- ϵ muitas vezes é um vetor de variáveis aleatórias
- O modelo de probabilidades é conhecido com precisão razoável
- Como escolher γ?

Risco esperado

Modelo probabilístico

- ϵ muitas vezes é um vetor de variáveis aleatórias
- O modelo de probabilidades é conhecido com precisão razoável
- Como escolher γ?

$$\gamma^* = \arg\min_{\gamma \in [0,\infty)} E_\epsilon \| B_\gamma(\boldsymbol{b} + \epsilon) - \boldsymbol{x} \|^2$$

Generalizando divergências

Divergência de Bregman

Generalizando divergências

Divergência de Bregman

- $f : \mathbb{R}^n \to \mathbb{R}$ estritamente convexa

Generalizando divergências

Divergência de Bregman

- $f : \mathbb{R}^n \to \mathbb{R}$ estritamente convexa
- $D_f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}) - f(\boldsymbol{y}) - \nabla f(\boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{y}) \geq 0$

Generalizando divergências

Divergência de Bregman

- $f : \mathbb{R}^n \to \mathbb{R}$ estritamente convexa
- $D_f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}) - f(\boldsymbol{y}) - \nabla f(\boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{y}) \geq 0$
- $D_{\|\cdot\|^2}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$

Generalizando divergências

Divergência de Bregman

- $f : \mathbb{R}^n \to \mathbb{R}$ estritamente convexa
- $D_f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}) - f(\boldsymbol{y}) - \nabla f(\boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{y}) \geq 0$
- $D_{\|\cdot\|^2}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$
- Como escolher γ?

Generalizando divergências

Divergência de Bregman

- $f : \mathbb{R}^n \to \mathbb{R}$ estritamente convexa
- $D_f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}) - f(\boldsymbol{y}) - \nabla f(\boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{y}) \geq 0$
- $D_{\|\cdot\|^2}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$
- Como escolher γ?

$$\gamma^* = \underset{\gamma \in [0,\infty)}{\arg\min}\, E_\epsilon D_f(\boldsymbol{x}, B_\gamma(\boldsymbol{b} + \epsilon))$$

Construindo substitutos para o erro esperado

Risco preditivo esperado de Bregman (RPEB)

Construindo substitutos para o erro esperado

Risco preditivo esperado de Bregman (RPEB)

- $\boldsymbol{x}_\gamma(\boldsymbol{b}) := B_\gamma(\boldsymbol{b})$

Construindo substitutos para o erro esperado

Risco preditivo esperado de Bregman (RPEB)

- $\boldsymbol{x}_\gamma(\boldsymbol{b}) := B_\gamma(\boldsymbol{b})$
- $R_\gamma^f(\boldsymbol{b}) := E_\epsilon D_f(A\boldsymbol{x}, A\boldsymbol{x}_\gamma(\boldsymbol{b} + \epsilon))$

Construindo substitutos para o erro esperado

Risco preditivo esperado de Bregman (RPEB)

- $\boldsymbol{x}_\gamma(\boldsymbol{b}) := B_\gamma(\boldsymbol{b})$
- $R_\gamma^f(\boldsymbol{b}) := E_\epsilon D_f(A\boldsymbol{x}, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\epsilon))$
- Como escolher γ?

Construindo substitutos para o erro esperado

Risco preditivo esperado de Bregman (RPEB)

- $\boldsymbol{x}_\gamma(\boldsymbol{b}) := B_\gamma(\boldsymbol{b})$
- $R^f_\gamma(\boldsymbol{b}) := E_\epsilon D_f(A\boldsymbol{x}, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\epsilon))$
- Como escolher γ?

$$\gamma^* = \arg\min_{\gamma\in[0,\infty)} R^f_\gamma(\boldsymbol{b})$$

Construindo substitutos para o erro esperado

Risco preditivo esperado de Bregman (RPEB)

- $\boldsymbol{x}_\gamma(\boldsymbol{b}) := B_\gamma(\boldsymbol{b})$
- $R^f_\gamma(\boldsymbol{b}) := E_\epsilon D_f(A\boldsymbol{x}, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\epsilon))$
- Como escolher γ?

$$\gamma^* = \arg\min_{\gamma\in[0,\infty)} R^f_\gamma(\boldsymbol{b})$$

$$\begin{aligned} R^f_\gamma(\boldsymbol{b}) &= E_\epsilon D_f(A\boldsymbol{x}, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\epsilon)) \\ &\neq E_\epsilon D_f(\boldsymbol{x}, \boldsymbol{x}_\gamma(\boldsymbol{b}+\epsilon)) \end{aligned}$$

Construindo substitutos para o erro esperado

Dificuldades

Construindo substitutos para o erro esperado

Dificuldades

- A "realidade" $\boldsymbol{x}$ não é conhecida

Construindo substitutos para o erro esperado

Dificuldades

- A “realidade” $\boldsymbol{x}$ não é conhecida
- Calcular o valor esperado não seria possível

Construindo substitutos para o erro esperado

Soluções

Construindo substitutos para o erro esperado

Soluções

- Lema de Stein

Construindo substitutos para o erro esperado

Soluções

- Lema de Stein
- Utilizar uma realização ao invés do valor esperado

Tornando o RPEB computável

Lema de Stein

Tornando o RPEB computável

Lema de Stein

Seja $\boldsymbol{b} \in \mathbb{R}^m$ e $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 I)$ e considere $\boldsymbol{h} : \mathbb{R}^m \mapsto \mathbb{R}^m$ fracamente diferenciável e tal que, para $i \in \{1, 2, \ldots, m\}$, $E_{\epsilon} \left| \frac{\partial h_i}{\partial b_i}(\boldsymbol{b} + \boldsymbol{\epsilon}) \right| < \infty$. Então

$$E_{\epsilon} \left[\boldsymbol{h}(\boldsymbol{b} + \boldsymbol{\epsilon})^T \boldsymbol{\epsilon} \right] = \sigma^2 E_{\epsilon} \left[\sum_{i=1}^{m} \frac{\partial h_i}{\partial b_i}(\boldsymbol{b} + \boldsymbol{\epsilon}) \right]. \tag{1}$$

Tornando o RPEB computável

Lema de Stein

Tornando o RPEB computável

Lema de Stein

- Permite trocar estimador desconhecido por computável

Tornando o RPEB computável

Lema de Stein

- Permite trocar estimador desconhecido por computável
- Existem variantes para diversos modelos de ruído

Tornando o RPEB computável

Lema de Stein

- Permite trocar estimador desconhecido por computável
- Existem variantes para diversos modelos de ruído
- Hipóteses bastante gerais

Tornando o RPEB computável

Lema de Stein

- Permite trocar estimador desconhecido por computável
- Existem variantes para diversos modelos de ruído
- Hipóteses bastante gerais
- Envolve a derivada de de uma função geralmente fornecida implicitamente

Tornando o RPEB computável

Usando o Lema de Stein

Tornando o RPEB computável

Usando o Lema de Stein

$$
\begin{aligned}
D_f\left(A\mathbf{x}, A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right) &= D_f\left(\mathbf{b}, A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right) \\
&= f(\mathbf{b}) - f\left(A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right) - \nabla f\left(A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right)^T\left(\mathbf{b} - A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right) \\
&= f(\mathbf{b}) - f\left(A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right) - \nabla f\left(A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right)^T\left(\mathbf{b} + \boldsymbol{\epsilon} - A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right) \\
&\qquad + \nabla f\left(A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right)^T\boldsymbol{\epsilon} \\
&= f(\mathbf{b}) - f(\mathbf{b}+\boldsymbol{\epsilon}) + D_f\left(\mathbf{b}+\epsilon, A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right) + \nabla f\left(A\mathbf{x}_\gamma(\mathbf{b}+\boldsymbol{\epsilon})\right)^T\boldsymbol{\epsilon}.
\end{aligned}
$$

Tornando o RPEB computável

Usando o Lema de Stein

Tornando o RPEB computável

Usando o Lema de Stein

$$
\begin{aligned}
R_\gamma^f(\boldsymbol{b}) &= K + E_\epsilon\left[D_f\left(\boldsymbol{b}+\epsilon, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right)\right] + E_\epsilon\left[\nabla f\left(A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right)^T\boldsymbol{\epsilon}\right] \\
&= K + E_\epsilon\left[D_f\left(\boldsymbol{b}+\epsilon, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right)\right] + E_\epsilon\left[\boldsymbol{h}(\boldsymbol{b}+\boldsymbol{\epsilon})^T\boldsymbol{\epsilon}\right] \\
&= K + E_\epsilon\left[D_f\left(\boldsymbol{b}+\epsilon, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right)\right] + \sigma^2 E_\epsilon\left[\sum_{i=1}^m \frac{\partial h_i}{\partial b_i}(\boldsymbol{b}+\boldsymbol{\epsilon})\right]
\end{aligned}
$$

Tornando o RPEB computável

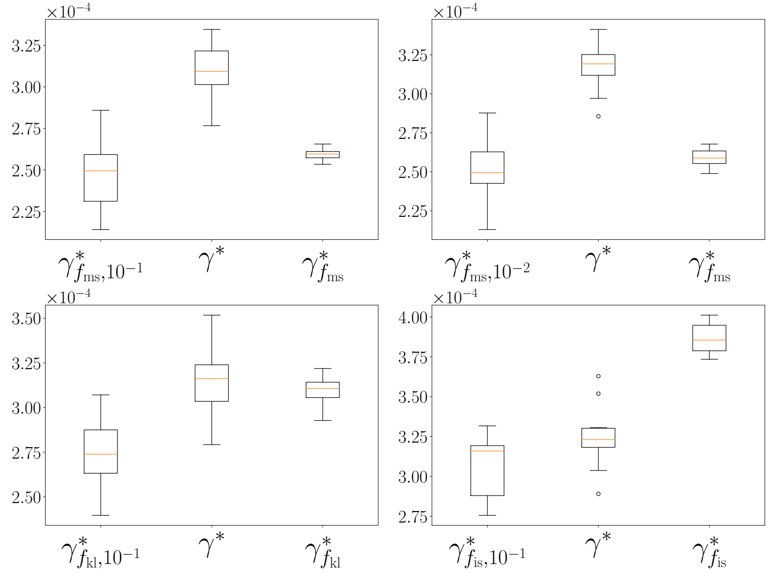
Usando o Lema de Stein

$$
\begin{aligned}
R_\gamma^f(\boldsymbol{b}) &= K + E_\epsilon\left[D_f\left(\boldsymbol{b}+\epsilon, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right)\right] + E_\epsilon\left[\nabla f\left(A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right)^T\boldsymbol{\epsilon}\right] \\
&= K + E_\epsilon\left[D_f\left(\boldsymbol{b}+\epsilon, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right)\right] + E_\epsilon\left[\boldsymbol{h}(\boldsymbol{b}+\boldsymbol{\epsilon})^T\boldsymbol{\epsilon}\right] \\
&= K + E_\epsilon\left[D_f\left(\boldsymbol{b}+\epsilon, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right)\right] + \sigma^2 E_\epsilon\left[\sum_{i=1}^m \frac{\partial h_i}{\partial b_i}(\boldsymbol{b}+\boldsymbol{\epsilon})\right]
\end{aligned}
$$

Portanto, podemos tentar usar o parâmetro γ que minimiza

$$
\begin{aligned}
\tilde{R}_\gamma^f(\boldsymbol{b}) &= D_f\left(\boldsymbol{b}+\epsilon, A\boldsymbol{x}_\gamma(\boldsymbol{b}+\boldsymbol{\epsilon})\right) + \sigma^2\sum_{i=1}^m \frac{\partial h_i}{\partial b_i}(\boldsymbol{b}+\boldsymbol{\epsilon}) \\
&\approx R_\gamma^f(\boldsymbol{b}) - K
\end{aligned}
$$

Experimento simulado



Conclusões

- Minimizar um risco preditivo de Bregman qualquer não é mais difícil do que minizar o risco preditivo quadrático

Conclusões

- Minimizar um risco preditivo de Bregman qualquer não é mais difícil do que minizar o risco preditivo quadrático
- Em alguns casos, a precisão pode ser melhor para outros riscos preditivos de Bregman do que para o risco preditivo quadrático

Conclusões

- Minimizar um risco preditivo de Bregman qualquer não é mais difícil do que minizar o risco preditivo quadrático
- Em alguns casos, a precisão pode ser melhor para outros riscos preditivos de Bregman do que para o risco preditivo quadrático
- Mesmo usando uma única realização a precisão é grande

Conclusões

- Minimizar um risco preditivo de Bregman qualquer não é mais difícil do que minizar o risco preditivo quadrático
- Em alguns casos, a precisão pode ser melhor para outros riscos preditivos de Bregman do que para o risco preditivo quadrático
- Mesmo usando uma única realização a precisão é grande
- Podemos mostrar que o resultado concentra-se fortemente em torno do valor esperado?