

3

Filtro, Previsor e Suavizador de Kalman

3.1

As Três Formas Básicas de Estimação do Vetor de Estado

No decorrer do desenvolvimento histórico da teoria dos modelos em EE, tanto sob o enfoque da Estatística, quanto sob o enfoque da Engenharia, ficou convencionalizado que a estimação do vetor de estado α_t de um dado modelo deve ser estratificada em três categorias básicas, conforme o tipo de informação – proveniente do processo estocástico das medidas Y_t – que estará sendo usado.

No intuito de que se organizem as especulações que se seguirão, considere-se, para tanto, um processo estocástico p-variado Y_t , composto de medidas, seguindo um modelo em EE do tipo descrito na Definição 2.1. Defina-se o vetor aleatório a_{tj} com as *projeções ortogonais* – ou, de maneira mais formal, com as *imagens da projeção ortogonal* – das coordenadas **univariadas** do vetor de estado α_t no subespaço fechado gerado pelas coordenadas **univariadas** de todas as medidas até o instante j ¹. Defina-se a matriz $P_{tj} \equiv E[(\alpha_t - a_{tj})(\alpha_t - a_{tj})']$ com os erros médios quadráticos *incondicionais* (ou *condicional* para o caso de o modelo em EE tiver pelo menos uma de suas matrizes do sistema dependente do passado das medidas) associados às projeções em a_{tj} . Ancoradas nesta nomenclatura, as três categorias de estimação podem ser enunciadas da seguinte forma:

- Se $j < t$, tem-se um problema de *previsão* ou *predição*;
- Se $j = t$, tem-se um problema de *filtragem* ou *atualização*; e
- Se $j > t$, tem-se um problema de *suavização* ou *interpolação*.

¹ Pode ser provado que o espaço das (classes de equivalências de) variáveis aleatórias de segunda ordem definidas em um apropriado espaço de probabilidade (Ω, Λ, P) , munido do produto interno usual dado pela esperança do produto cruzado, é um espaço de *Hilbert*. Logo, existe tal projeção – transformação linear idempotente com núcleo ortogonal à imagem – e esta é única, de acordo com o Teorema 5.52 de Kubrusly (2001). No próximo capítulo, estes tópicos serão novamente revisitados.

O que as recursões do Filtro de Kalman fornecem são justamente as projeções ortogonais das coordenadas dos vetores de estado sobre o subespaço gerado por um conjunto de informações provenientes de medidas, as quais são dadas convenientemente em uma forma analítica que explora, como será visto mais adiante, a dinâmica temporal do modelo em EE linear. De acordo com o tipo de informação utilizado, dá-se às recursões o nome *previsor*, *filtro* ou *suavizador* (ou, de forma equivalente, *preditor*, *atualizador* ou *interpolador*), e ao vetor aleatório $a_{t/j}$ o nome de *previsto*, *filtrado* ou *suavizado* (ou, de forma equivalente, *predito*, *atualizado* ou *interpolado*).

Antes de se apresentarem as recursões de Kalman para cada uma das três categorias, é preciso e conveniente que se estabeleça uma notação básica, qual seja:

- $a_{t/j}$ e $P_{t/j}$ são definidos como antes.
- $a_t \equiv a_{t/t-1}$ é o vetor de previstos 1 passo à frente;
- $P_t \equiv P_{t/t-1} = E[(\alpha_t - a_t)(\alpha_t - a_t)']$;
- $v_t \equiv Y_t - \hat{Y}_t = Y_t - Z_t a_t - d_t$ é o vetor de *inovações*; e
- $F_t \equiv E(v_t v_t') = Z_t P_t Z_t' + H_t$.

3.2

Previsão

Para o problema de previsão, as recursões de Kalman são dadas de forma bem natural, a qual explora as propriedades de linearidade da projeção ortogonal. Suponha-se, então, que seja de interesse prever o vetor de estado no instante $t + h$, para algum horizonte h , com base na informação proveniente das medidas até t . Na notação estabelecida na seção anterior, deseja-se obter $a_{t+h/t}$. Usando a estrutura da equação (2.1), deduz-se que as recursões desejadas² são

² Supõe-se, sem perda de generalidade, que as matrizes do sistema não dependem do passado das medidas. Caso contrário, algumas modificações rigorosas são necessárias para $h > 1$, as quais não são exploradas nesta Dissertação.

$$\begin{aligned} \mathbf{a}_{t+h/t} &= \mathbf{T}_{t+h-1} \mathbf{a}_{t+h-1/t} + \mathbf{c}_{t+h-1} \\ \mathbf{P}_{t+h/t} &= \mathbf{T}_{t+h-1} \mathbf{P}_{t+h-1/t} \mathbf{T}'_{t+h-1} + \mathbf{R}_{t+h-1} \mathbf{Q}_{t+h-1} \mathbf{R}'_{t+h-1} \end{aligned} \quad (3.1)$$

Um caso particular de grande interesse é a questão da previsão 1 passo à frente. Para tanto, basta fazer $h = 1$ nas recursões em 3.1, o que gera as comumente chamadas *equações de previsão do Filtro de Kalman* – nomenclatura que será usada no decorrer da Dissertação – as quais são dadas por

$$\begin{aligned} \mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_{t/t} + \mathbf{c}_t \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_{t/t} \mathbf{T}'_t + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}'_t \end{aligned} \quad (3.2)$$

As equações em (3.2) desempenham papel importante, pois estas são necessárias para a implementação das recursões referentes ao problema de atualização a ser discutido a seguir.

3.3

Atualização

Para o problema de atualização, as respectivas recursões do Filtro de Kalman são dados de acordo com o seguinte

Teorema 3.1. *(O Filtro de Kalman) Seja Y_t um processo estocástico p -variado com uma representação em EE do tipo dado nas equações em (2.1). Então, o vetor de projeções $\mathbf{a}_{t/t}$ e a matriz de erros médios quadráticos associada $\mathbf{P}_{t/t}$ são dados por*

$$\begin{aligned} \mathbf{a}_{t/t} &= \mathbf{a}_t + \mathbf{P}_t \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{v}_t \\ \mathbf{P}_{t/t} &= \mathbf{P}_t - \mathbf{P}_t \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_t \end{aligned} \quad (3.3)$$

Se, porventura, \mathbf{F}_t não for invertível, então se pode fazer uso de uma inversa generalizada.

Prova: Vide Brockwell e Davis (1996) ou Duncan e Horn (1972). \square

São pertinentes as seguintes observações a respeito do último Teorema:

- 1) As recursões em (3.3) são chamadas na literatura de *equações de atualização do Filtro de Kalman*. Esta nomenclatura será também adotada nesta Dissertação.
- 2) Como pode ser visto, as equações de atualização dependem das equações de previsão, e vice-versa. Esta característica promove um meio de se suprimir uma delas via processo de substituição. Por exemplo, as expressões das equações de atualização podem ser substituídas nas expressões das equações de previsão, gerando o conhecido *Filtro de Kalman “2 em 1”*³, cujas expressões finais são dadas por

$$\begin{aligned} a_{t+1} &= T_t a_t + c_t + T_t P_t Z_t' F_t^{-1} v_t \\ P_{t+1} &= T_t P_t (T_t - T_t P_t Z_t' F_t^{-1} Z_t)' + R_t Q_t R_t' \end{aligned} \quad (3.4)$$

A substituição inversa (equações de previsão nas equações de atualização), pouco demandada em situações práticas, também pode ser obtida. Para detalhes e expressão final, vide, por exemplo, Bertsekas (1995), página 364.

- 3) Quando a matriz F_t não é invertível, o que se tem, verdadeiramente, é uma dependência linear entre as coordenadas univariadas das medidas até o instante t . No entanto, independentemente da inversa generalizada escolhida, a projeção obtida ainda mantém sua propriedade de unicidade.
- 4) Pode ser provado (vide seção 3.5) que o estimador $a_{t/t}$ é mais preciso do que o estimador a_t , no sentido de que $P_{t/t} \leq P_t$ na ordenação usual das matrizes simétricas.
- 5) Cita-se que estas recursões produzem estimadores *lineares* ótimos sob função de perda quadrática. Se, por ventura, o modelo em EE linear for Gaussiano, então se pode provar – vide, por exemplo, Harvey (1989), Harvey (1993), Tanizaki (1996) ou Durbin e Koopman (2001a) – que tais recursões constituem em *esperanças condicionais* e, portanto, são

³ Embora esta nomenclatura seja eventualmente usada na Dissertação, ela não necessariamente é adotada de forma padrão na literatura.

estimadores ótimos *globais*, ainda sob a função de perda quadrática. Este mesmo comentário se estende para o Suavizador de Kalman, o qual é apresentado na próxima seção.

3.4

Suavização

É examinada agora a questão da suavização, ou seja, como são os estimadores do vetor de estado em um dado instante de tempo, com base em informações geradas pelas medidas até este mesmo instante juntamente com a informação proveniente do “futuro” (baseada em medidas de instantes de tempo posteriores).

O problema da suavização pode ser, por sua vez, subdividido em três enfoques, quais sejam:

- *Ponto fixo*;
- *Defasagem fixa*; e
- *Intervalo Fixo*.

Os dois primeiros enfoques são procedimentos *on-line* e podem ser estudados, por exemplo, em Harvey (1989) e, com maior riqueza de detalhes, por exemplo, em Anderson e Moore (1979).

O *enfoque do intervalo fixo*, por sua vez, consiste em um procedimento *off-line*, o qual, como investigado mais adiante, se constitui de recursões do tipo *backward*, isto é, que “andam de trás para frente”, percorrendo desde o “futuro” até o vetor de estado inicial. Este enfoque é o preferencial para análises estatísticas, idéia compartilhada também por Harvey (1989) e por Tanizaki (1996).

Historicamente, o suavizador do intervalo fixo de Kalman passou por duas formulações matematicamente equivalentes, porém de deduções e eficiências computacionais completamente diferentes. A primeira delas é alicerçada em uma dedução mais difícil, em um número maior de multiplicações e de inversões matriciais, e demanda armazenamento de mais estruturas matriciais do que a

segunda. Portanto, não é difícil concluir, inclusive com base em publicações mais recentes sobre o assunto, que esta primeira formulação está caindo paulatinamente em desuso. Existem diversas referências que apresentam sua dedução, sendo três das mais harmonizadas ao contexto estatístico as de Ansley e Kohn (1982), Kohn e Ansley (1983) e Hamilton (1994).

A segunda formulação, sob outra visada, apresenta diversos atrativos, como citado em de Jong (1989), o qual é, provavelmente, o artigo mais abrangente e concertado sobre o tema dentro do contexto estatístico: tem solução analítica, possui eficiência computacional considerável, abrange casos degenerados com facilidade etc.

Entretanto, antes de tudo, é conveniente estabelecer a seguinte notação. Suponha-se, a partir de agora, que seja extraída uma *série temporal* de tamanho n do processo de medidas Y_t . Denote-se $\hat{\alpha}_t \equiv a_{t/n}$ e $V_t \equiv P_{t/n}$. Com base nesta notação, apresenta-se agora o

Teorema 3.2. (O Suavizador do Intervalo Fixo de Kalman) *Seja Y_t um processo estocástico p -variado com uma representação em EE do tipo dado nas equações em (2.1). Então, o vetor $\hat{\alpha}_t$ e a matriz V_t são dados pelas expressões*

$$\begin{aligned}\hat{\alpha}_t &= a_t + P_t r_{t-1} \\ r_{t-1} &= Z_t' F_t^{-1} v_t + (T_t - T_t P_t Z_t' F_t^{-1} Z_t)' r_t \\ V_t &= P_t - P_t N_{t-1} P_t \\ N_{t-1} &= Z_t' F_t^{-1} Z_t + (T_t - T_t P_t Z_t' F_t^{-1} Z_t)' N_t (T_t - T_t P_t Z_t' F_t^{-1} Z_t)\end{aligned}\tag{3.5}$$

com $r_n = 0$ e $V_n = 0$. Se, porventura, F_t não for invertível, então se pode fazer uso de uma inversa generalizada.

Prova: Vide de Jong (1989). □

Podem ser feitas as seguintes observações sobre este último Teorema:

- 1) Observa-se nas recursões em (3.5) que as equações de previsão e de atualização de Kalman – ou, de forma compacta, o Filtro de Kalman

“2 em 1” – desempenham papel fundamental para a suavização de intervalo fixo. Mais especificamente, as quantidades a_t , P_t e F_t dadas diretamente pelas equações em (3.4) se constituem em um *input* básico, devendo ser armazenadas para o cálculo dos suavizados.

- 2) Pode ser provado (vide seção 3.5) que o estimador \hat{a}_t é mais preciso do que o estimador $a_{t/t}$, no sentido de que $V_t \leq P_{t/t}$ na ordenação usual das matrizes simétricas.
- 3) Constata-se, claramente, que *não há uma inversão de matriz adicional sequer* (F_t é invertida antes! – vide as equações de atualização em (3.3) ou as equações do filtro “2 em 1” em (3.4)) nesta formulação de de Jong, o que corrobora, superlativamente, sua eficiência computacional. A formulação original, no outro extremo, demanda, além de algumas multiplicações adicionais, mais $n-1$ inversões matriciais! Para confirmação deste grande aumento de operações matriciais, o pesquisador pode consultar as deduções de Ansley e Kohn (1982) ou, se preferir, apenas as expressões finais em Harvey (1989), página 154.
- 4) As quantidades r_t e N_t , além de serem usadas como quantidades intermediárias no Suavizador do Kalman, também fazem parte de um conjunto de técnicas de diagnósticos para o modelo em EE, como descrito em pormenores em de Jong e Penzer (1998) e de forma resumida no final do capítulo 7 de Durbin e Koopman (2001a). Também é motivador notar que N_t é a matriz de covariâncias do vetor aleatório r_t , afirmação devidamente demonstrada no capítulo 4 de Durbin e Koopman (2001a).

3.5

Hierarquia de Precisão para as Três Formas de Estimação

Enuncia-se e prova-se agora o resultado já previamente comentado sobre a ordenação existente das precisões relativas ao Previsor, ao Filtro e ao Suavizador de Kalman. Para tanto, tenha-se em mente o modelo em EE linear que vem sido discutido, juntamente com as recursões de estimação apresentadas nas seções anteriores.

Proposição 3.3. (Hierarquia de precisão) $V_t \leq P_{t|t} \leq P_t$, na ordenação usual das matrizes simétricas.

Prova: Seja $x \in \mathbb{R}^m$. Então,

$$\begin{aligned} x'V_t x &= x'(P_t - P_t N_{t-1} P_t)x \\ &= x'(P_t - P_t(Z_t' F_t^{-1} Z_t + (T_t - T_t P_t Z_t' F_t^{-1} Z_t)' N_t (T_t - T_t P_t Z_t' F_t^{-1} Z_t)) P_t)x \\ &\leq (N_t \text{ é não negativa definida}) \leq \\ &\leq x'(P_t - P_t Z_t' F_t^{-1} Z_t P_t)x = x'P_{t|t}x. \text{ Isto demonstra a primeira} \\ &\text{desigualdade.} \end{aligned}$$

Seguindo, tem-se que

$$x'P_{t|t}x = x'(P_t - P_t Z_t' F_t^{-1} Z_t P_t)x \leq (F_t^{-1} \text{ é não negativa definida}) \leq x'P_t x.$$

Isto demonstra a segunda desigualdade. \square

Esta última proposição tem um apelo intuitivo bastante claro, cujo significado pode ser resumido no seguinte: **quanto maior a informação usada pelo estimador, melhor será sua precisão.**

3.6

Inicialização do Filtro de Kalman

Para a implementação das recursões de Kalman apresentadas nas seções anteriores, é necessário, na maioria das situações práticas, que se conjecture um tratamento para o vetor de estado inicial α_1 . Na Definição 2.1, se apresenta a hipótese de que tal vetor aleatório deveria ser de segunda ordem e não correlacionado com os processos de erros da equação das medidas e da equação do estado. No entanto, eventualmente há de se relaxar o pressuposto de integralidade de segunda ordem, de forma que todas as componentes univariadas de α_1 , associadas a coordenadas univariadas não estacionárias do processo α_t , sejam *difusas*, isto é, suas variâncias sejam arbitrariamente grandes. Este tipo de procedimento se enquadra, de certa maneira, na questão de *adoção de prioris não informativas* em análises Bayesianas. Para uma discussão a respeito deste último

assunto – o de *prioris* não informativas – sugere-se aos interessados o estudo do capítulo 3 de Migon e Gamerman (1999).

No contexto de modelos em Espaço de Estado, esta abordagem carece de uma caracterização um tanto mais particular e mais apropriada.

Definição 3.4. *Dá-se o nome de inicialização difusa a qualquer tipo de inicialização do Filtro de Kalman para um modelo em EE que abranja, pelo menos, uma coordenada difusa no seu vetor de estado inicial.*

Salienta-se que a teoria e a implementação de inicializações difusas, embora indiscutivelmente importantes, se constituem, talvez, na parte mais laboriosa da modelagem em EE linear.

Durbin e Koopman (2001a) propõem o chamado *modelo geral* para o vetor de estado inicial, o qual é dado por

$$\alpha_1 = \underline{a} + A\delta + R_0\eta_0, \quad (3.6)$$

no qual \underline{a} é um vetor fixo $m \times 1$ e conhecido (geralmente nulo), δ é um vetor de dimensão q de quantidades desconhecidas e estocásticas com vetor de média nulo e matriz de covariâncias κI , η_0 é vetor aleatório de segunda ordem com vetor de médias nulo e matriz de covariâncias não negativa definida Q_0 ⁴, e as matrizes A e R_0 são constituídas de vetores colunas da matriz identidade $I_{m \times m}$, de forma que $A'R_0 = 0$. A motivação da inicialização difusa é fazer justamente com que κ tenda a infinito, representando ignorância por do analista sobre como é o comportamento estocástico do bloco δ , representante das componentes não estacionárias, do vetor de estado inicial.

Com base no modelo geral dado em (3.6), pode ser deduzido diretamente que

$$P_1 = \kappa P_\infty + P_*, \quad (3.7)$$

⁴ Em Durbin e Koopman (2001a), η_0 é dito ter distribuição normal e sua matriz de covariâncias Q_0 deve ser positiva definida. Nesta Dissertação, foi feita a generalização pertinente, de forma a concordar com o modelo em EE linear previamente definido no capítulo 2.

na qual $P_\infty = AA'$ e $P_* = R_0Q_0R_0'$. O ponto interessante nesta Teoria é que, uma vez que P_∞ não é nula – implicando na existência de fato de alguns termos não estacionários no modelo em EE – a mesma incerteza máxima sobre as componentes de δ pode se propagar para alguns estados subseqüentes (a distribuição *a posteriori* do estado não será própria imediatamente!), até um instante de tempo d , após o qual tudo se estabiliza – pois a distribuição *a posteriori* do estado se torna *própria*. Alternativamente, pode ser dito que os vetores de estado, até um instante d , são difusos, “por herança” de α_1 . Se este d não existir para uma dada série temporal, Durbin e Koopman (2001a) definem a situação em tela como *degenerada*, ou seja, não existe informação suficiente para o ajuste do modelo em EE proposto (em outras palavras, a série é “pequena”).

3.6.1

Inicialização Difusa Aproximada

Uma prática bastante usada para modelos em EE lineares com pelo menos uma coordenada não estacionária é aproximar κ por um número muito grande (em geral, escolhe-se um valor de κ entre 10^3 e 10^7), incorrendo, assim, em uma espécie de solução aproximada para o problema de inicialização difusa.

O atrativo desta abordagem é a facilidade de implementação. No entanto, se verificam dois inconvenientes:

- 1) Não se sabe ao certo quando é o instante de estabilização d . Algumas regras de caráter *ad hoc*, ou eventualmente dedutíveis de forma rigorosa, costumam ser usadas; por exemplo, em um modelo de tendência linear local (vide capítulo 2 de Harvey (1989) ou capítulo 3 de Durbin e Koopman (2001a) para melhores informações), adota-se $d = 3$ – isto é, perdem-se as duas primeiras observações. Mas não há, em geral, sustentações rigorosas de que um dado instante d escolhido é o correto ou não o é, principalmente para modelos multivariados (ou seja, quando Y_t for multidimensional).
- 2) Em Koopman (1997), há uma discussão empírica que compara este tipo de solução aproximada com uma *exata*, sendo esta última tópico da próxima

subseção. A conclusão lá encontrada é a de que as duas abordagens destoam entre si, indicando que a inicialização difusa aproximada gera erros que se propagam no decorrer da filtragem de Kalman.

Apesar destes obstáculos, a inicialização difusa aproximada continua a ser preferencial na maioria de análises práticas com modelagem em EE linear.

3.6.2

Inicialização Difusa Exata

Um dos temas de pesquisa no escopo da modelagem em EE, ainda relativamente recente, foi justamente a obtenção de soluções exatas para o problema de inicialização difusa do Filtro de Kalman. Talvez o artigo que mais tenha se destacado neste nicho seja Koopman (1997). Nesse texto, é proposto o chamado *Filtro de Kalman Exato*, o qual se baseia na exploração do modelo geral dado em (3.6) para o vetor de estado inicial e se caracteriza por um conjunto de recursões que coincidem com as usuais de Kalman no período em que o efeito da inicialização difusa (aqui, exata) já tenha se extinguido – isto é, de $d+1$ em diante. Dentre as vantagens desta metodologia, citam-se:

- 1) Fornece, sob enfoque teórico e acadêmico, uma solução bem mais convincente do que a inicialização difusa aproximada;
- 2) Elimina, por definição, qualquer propagação de erro de aproximação, uma vez que esta se constitui numa solução exata;
- 3) Fornece, de maneira elegante e automática, o instante de estabilização d , o qual “aparece” nas recursões. Salienta-se que também é possível, sob esta abordagem exata, garantir a existência de tal número natural d ;
- 4) Promove recursões apropriadas também para o Suavizador de Kalman, inclusive no período difuso (isto é, em $t = 1, \dots, d$); e
- 5) Promove uma forma direta de se construir a função de verossimilhança *exata* para a estimação de parâmetros desconhecidos referentes a modelos lineares (condicionalmente) Gaussianos (vide capítulo 7).

As recursões do Filtro de Kalman Exato não serão aqui apresentadas. Encoraja-se, contudo, o pesquisador interessado nesta questão a estudar as páginas de 101 a 109 de Durbin e Koopman (2001a) e, para obtenção de informações adicionais, o artigo de 1997 de Koopman já citado. Adverte-se que estes são textos relativamente intrincados, estando em um nível algébrico bem acima das outras questões concernentes à teoria dos modelos em EE lineares.

Uma outra maneira de tratamento exato para a questão da inicialização é o *Filtro de Kalman Aumentado*, o qual também é discutido em Durbin e Koopman (2001a). Esta tecnologia, no entanto, se mostra inferior ao Filtro de Kalman Exato, devido às suas maiores complexidade e demanda computacional. Especificidades sobre estas comparações estão em Durbin e Koopman (2001a).

Na parte aplicada da Dissertação (capítulo 9), *todos os modelos em EE lineares a serem propostos e ajustados receberão um tratamento de inicialização difusa exata, via o Filtro de Kalman Exato.*