

Стійке байєсівське фільтрування та згладжування з використанням розподілу Стюдента t

M. Roth¹, T. Ardeshiri^{1,2}, E. Özkan³, and F. Gustafsson¹

¹*Dept. of Electrical Engineering, Linköping University,
SE-581 83 Linköping, Sweden*

²*Dept. of Engineering, Cambridge University, Cambridge CB2 1PZ, UK*

³*Dept. of Electrical and Electronics Engineering, Middle East Technical
University, Ankara, 06531 Turkey*

michael.roth@liu.se

Анотація

Оцінювання стану в умовах процесного й вимірювального шуму з важкими “хвостами” (heavy-tailed) — важливий виклик, який треба розв’язувати, наприклад, у сценаріях супроводу маневрених цілей з вимірюваннями, забрудненими “викидами” (outliers). Продуктивність фільтра Калмана (KF) у таких застосуваннях може погіршуватися через його тісний зв’язок із гаусівським розподілом. Тому ця стаття описує використання розподілу Стюдента t для розробки стійких, масштабованих і простих алгоритмів фільтрування та згладжування.

Після обговорення розподілу Стюдента t аналізується точне фільтрування в лінійних просторово-станових моделях з t -шумом. Проміжні кроки наближення використовуються для отримання алгоритмів фільтрування та згладжування, які тісно нагадують KF і згладжувач Рауха-Тунґа-Штрібеля (RTS), за винятком нелінійного матричного оновлення, що залежить від вимірювань. Обговорюються необхідні наближення і виявляється небажана поведінка узгодження моментів (moment matching) для t -щільностей. Представлено сприятливе наближення на основі мінімізації розбіжності Кульбака-Лейблера. Завдяки зв’язку з KF t -фільтр успадковує деякі властивості та алгоритмічні розширення. Показові симуляційні приклади демонструють продуктивність і стійкість нових алгоритмів.

1. Вступ

Фільтр Калмана (KF) — основний інструмент для оцінювання в лінійних просторово-станових моделях. Його властивості оптимальності як найкращого лінійного фільтра у сенсі мінімальної дисперсії добре встановлено [?]. Його виведення як оптимального байєсівського фільтра для гаусівського шуму [?] робить його простим для розуміння. Однак тісний зв'язок з гаусівським розподілом також тягне певні виклики, бо багато явищ реального світу не можна добре описати гаусівським розподілом. Приклади включають: викиди вимірювань, що їх продукують ненадійні сенсори; маневри цілі, які можна сприймати як стрибок процесного шуму; помилки лінеаризації у наближених нелінійних моделях. Прагматичний спосіб підійти до таких викликів — припустити процесний і вимірювальний шум з важкими хвостами. Тому ми досліджуємо розподіл Стюдента t як “родича” гаусівського розподілу з важкими хвостами та його використання для фільтрування і згладжування.

Ця стаття подає прості алгоритми оцінювання стану в дусі KF. Наш підхід — дослідити байєсівські рекурсії фільтрування та згладжування [?] для лінійних систем з t -шумом Стюдента. За допомогою спільних апроксимацій t -щільностей, результати для t -розподілу далі використовуються для отримання зручних оновлень за часом і вимірюванням. Отриманий фільтр виглядає схоже до KF, за винятком матричного оновлення, що нелінійно залежить від вимірювання y_k , та проміжного масштабування матричних параметрів. Згладжувач еквівалентний оберненому проходу Рауха-Тунга-Штрібеля [?, ?]. Окрім розробки алгоритму, стаття подає обговорення розподілу Стюдента t і важливих відмінностей, які треба враховувати при заміні гаусівського шуму на t -шум. Кроки апроксимації фільтра проаналізовано і виведено оптимальні вибори параметрів за критерієм розбіжності Кульбака-Лейблера. Крім того, аналіз еліптично-контурних розподілів показує походження поширених KF-виразів.

t -фільтр з цієї статті вперше запропонували в [?, ?]. Дана стаття заповнює багато деталей, що залишились відкритими в [?] — як-от: як найкраще виконувати проміжні кроки апроксимації та потенційні проблеми з узгодженням моментів. Крім того, ми доповнюємо фільтр алгоритмом згладжування. Обговорення еліптично-контурних розподілів та аналіз кроку точного фільтрування для шуму Стюдента t можуть слугувати основою для розробки алгоритмів поза нашими запропонованими рішеннями. Важливо, що ми вважаємо цю роботу сходинкою у розвитку складніших алгоритмів. Тому також обговорено розширення для застосування у нелінійних моделях.

Споріднені роботи з фільтрування у шумі з важкими хвостами включають: [?, ?] як ранні роботи з викидів вимірювань; [?, ?, ?] як ранні підходи на основі Стюдентового та еліптично-контурних розподілів, але без послідовних апроксимацій; [?, ?] та [?] як варіаційні підходи; [?] як оптимізаційний підхід; [?, ?] як фільтри на основі важкохвостого розподілу Леві; [?] як фільтр частинок для супроводу маневрених цілей; [?, ?] як сигма-точкові варіанти оригінального t -фільтра [?] для нелінійних систем. Згладжування: [?, ?] як варіаційні підходи; [?] як батч-KF на основі стійкого оцінювання максимальної правдоподібності; [?] як родич нашого згладжувача, але з повторюваним використанням узгодження моментів.

Структуру статті описано далі. Розподіл Стюдента t та сім'я еліптично-контурних розподілів обговорюються у розділі 2. Байєсівське фільтрування та згладжування за наявності t -шуму обговорюється у розділі 3. Алгоритм фільтра Стюдента t вводиться у розділі 4, а деякі його властивості виокремлюються в ???. Згладжувач Стюдента t — у розділі ???. Симуляційні приклади подано в 5 та підсумовано у заключних зауваженнях 6.

2. Розподіл Стюдента t : ключові результати

n -вимірний випадковий вектор ξ , що має розподіл Стюдента t , характеризується вектором середнього μ , додатно (напів)визначеною симетричною матрицею масштабу Σ розміру $n \times n$, та скалярним числом ступенів свободи $\nu > 0$. Менші значення ν дають важчі хвости. Для додатно визначеної Σ функція густини має вигляд

$$\text{St}(\xi; \mu, \Sigma, \nu) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(\nu\pi)^{n/2}} \frac{1}{\sqrt{\det(\Sigma)}} \left(1 + \frac{1}{\nu}(\xi - \mu)^T \Sigma^{-1}(\xi - \mu)\right)^{-\frac{\nu+n}{2}}. \quad (1)$$

Альтернативно, густину можна записати як нескінченну гаусівську суміш [?] зі скритою (latent) гамма-розподіленою змінною λ :

$$\text{St}(\xi; \mu, \Sigma, \nu) = \int \mathcal{N}(\xi; \mu, \frac{1}{\lambda}\Sigma) \mathcal{G}(\lambda; \frac{\nu}{2}, \frac{\nu}{2}) d\lambda. \quad (2)$$

При $\nu \rightarrow \infty$ гамма-щільність прямує до дельта-функції в 1, і $\text{St}(x; \mu, \Sigma, \nu)$ збігається до $\mathcal{N}(x; \mu, \Sigma)$.

Коваріаційна матриця ξ скінченна лише при $\nu > 2$ і задається як

$$\text{cov}(\xi) = \frac{\nu}{\nu-2}\Sigma, \quad \nu > 2. \quad (3)$$

При $\nu = 1$ маємо розподіл Коші, який не має середнього.

Лінійні перетворення t -векторів зберігають ступені свободи. Середнє і матриця масштабу перетворюються аналогічно як у гаусівському випадку. Для розбитих векторів ξ з

$$p(\xi) = p(\xi_1, \xi_2) = \text{St} \left(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}; \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_2 \end{pmatrix}, \nu \right), \quad (4)$$

маргінальна щільність ξ_2 :

$$p(\xi_2) = \text{St}(\xi_2; \mu_2, \Sigma_2, \nu). \quad (5)$$

Умовна щільність ξ_1 за ξ_2 — також t -щільність, але з більшими ступенями свободи:

$$p(\xi_1 | \xi_2) = \text{St}(\xi_1; \mu_{1|2}, \Sigma'_{1|2}, \nu_{1|2}), \quad (6)$$

з параметрами

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_2^{-1}(\xi_2 - \mu_2) = \mu_1 + \Upsilon(\xi_2 - \mu_2), \quad (7a)$$

$$\Sigma_{1|2} = \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^T = \Sigma_1 - \Upsilon\Sigma_2\Upsilon^T, \quad (7b)$$

$$\Sigma'_{1|2} = \frac{\nu + (\xi_2 - \mu_2)\Sigma_2^{-1}(\xi_2 - \mu_2)^T}{\nu + n_2} \Sigma_{1|2}, \quad (7c)$$

$$\nu_{1|2} = \nu + n_2, \quad (7d)$$

де введено “матрицю підсилення” $\Upsilon = \Sigma_{12}\Sigma_2^{-1}$. Умовне середнє таке саме, як у гаусівському випадку. Умовна матриця масштабу збігається з гаусівською, але масштабована множителем, що *нелінійно* залежить від ξ_2 . Уважний читач упізнає тут зв’язок із кроком оновлення вимірюваннями фільтра Калмана.

3. Постановка байєсівського фільтрування

Розглядаємо байєсівське фільтрування для лінійних просторово-станових моделей

$$x_{k+1} = Fx_k + v_k, \quad (8a)$$

$$y_k = Hx_k + e_k, \quad (8b)$$

де x_k — n -вимірний стан і y_k — m -вимірне вимірювання в момент k . Початковий стан x_0 , процесний і вимірювальний шум v_k, e_k є взаємно незалежними та розподіленими за Стьюдентом t :

$$p(x_0) = \text{St}(x_0; \hat{x}_0, P_0, \eta_0), \quad (9a)$$

$$p(v_k) = \text{St}(v_k; 0, Q, \gamma), \quad (9b)$$

$$p(e_k) = \text{St}(e_k; 0, R, \delta). \quad (9c)$$

Для білого шуму це Марковська модель з перехідною щільністю та правдоподібністю

$$p(x_{k+1} | x_k) = \text{St}(x_{k+1}; Fx_k, Q, \gamma), \quad (10a)$$

$$p(y_k | x_k) = \text{St}(y_k; Hx_k, R, \delta). \quad (10b)$$

Байєсівське оцінювання стану зводиться до знаходження умовних щільностей $p(x_k | y_{1:L})$. Розрізняють: $k > L$ — прогнозування, $k = L$ — фільтрування, $k < L$ — згладжування.

4. Алгоритм t -фільтра Стюдента

4.1. Простий фільтр на основі проміжних апроксимацій

Подібно до точного випадку, відправною точкою є t -щільність $p(x_k | y_{1:k}) = \text{St}(x_k; \hat{x}_{k|k}, P_{k|k}, \eta_k)$.

Перший виклик в одно-кроковому прогнозуванні — те, що проміжна спільна щільність $p(x_k, x_{k+1} | y_{1:k})$ містить добуток t -щільностей. Якщо, проте, ми пожертвуємо незалежністю і припустимо спільну t -щільність

$$p(x_k, v_k | y_{1:k}) = \text{St} \left(\begin{pmatrix} x_k \\ v_k \end{pmatrix}; \begin{pmatrix} \hat{x}_{k|k} \\ 0 \end{pmatrix}, \begin{pmatrix} P'_{k|k} & 0 \\ 0 & Q' \end{pmatrix}, \eta'_k \right) \quad (11)$$

зі сполученими ступенями свободи η'_k та параметрами $P'_{k|k}$ і Q' , то

$$p(x_k, x_{k+1} | y_{1:k}) = \text{St} \left(\begin{pmatrix} x_k \\ x_{k+1} \end{pmatrix}; \begin{pmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \end{pmatrix}, \begin{pmatrix} P'_{k|k} & P'_{k|k} F^T \\ F P'_{k|k} & P_{k+1|k} \end{pmatrix}, \eta'_k \right) \quad (12)$$

впливає за правилами лінійних перетворень t -векторів. Щільність прогнозу $p(x_{k+1} | y_{1:k}) = \text{St}(x_{k+1}; \hat{x}_{k+1|k}, P_{k+1|k}, \eta'_k)$ з параметрами

$$\hat{x}_{k+1|k} = F \hat{x}_{k|k}, \quad (13a)$$

$$P_{k+1|k} = F P'_{k|k} F^T + Q', \quad (13b)$$

впливає негайно. Параметри прогнозу нагадують часове оновлення KF. Однак $P_{k+1|k}$ тут інтерпретується як матриця масштабу, а не коваріаційна матриця.

Один з варіантів — узяти $\eta'_k = \min(\eta_k, \gamma)$, щоб зберегти найважчі хвости серед апостеріорних x_k і v_k . Вибір $Q' = Q$ і $P'_{k|k} = P_{k|k}$ — консервативний у сенсі, що приймається маргінальна коваріація більша за оригінальну.

Для підготовки оновлення за вимірюванням ми мусимо поєднати прогноз $p(x_k | y_{1:k-1}) = \text{St}(x_k; \hat{x}_{k|k-1}, P_{k|k-1}, \eta'_{k-1})$ з $p(e_k)$. Подібна спільна t -апроксимація:

$$p(x_k, e_k | y_{1:k-1}) = \text{St} \left(\begin{pmatrix} x_k \\ e_k \end{pmatrix}; \begin{pmatrix} \hat{x}_{k|k-1} \\ 0 \end{pmatrix}, \begin{pmatrix} P'_{k|k-1} & 0 \\ 0 & R' \end{pmatrix}, \eta''_k \right), \quad (14)$$

зі сполученими ступенями свободи η''_k та підкоригованими матрицями $P'_{k|k-1}$ і R' . Простий вибір: $\eta''_k = \min(\eta'_k, \delta)$, $P'_{k|k-1} = P_{k|k-1}$, $R' = R$. Спільна щільність стану та виходу:

$$p(x_k, y_k | y_{1:k-1}) = \text{St} \left(\begin{pmatrix} x_k \\ y_k \end{pmatrix}; \begin{pmatrix} \hat{x}_{k|k-1} \\ \hat{y}_k \end{pmatrix}, \begin{pmatrix} P'_{k|k-1} & P'_{k|k-1} H^T \\ H P'_{k|k-1} & S_k \end{pmatrix}, \eta''_k \right). \quad (15)$$

Прогноз виходу та його коваріація (аналог S у KF):

$$\hat{y}_k = H \hat{x}_{k|k-1}, \quad (16a)$$

$$S_k = H P'_{k|k-1} H^T + R'. \quad (16b)$$

Нарешті, оновлення за вимірюванням впливає з умовної t -щільності (6). Аналогічно KF, матриця підсилення

$$K_k = P'_{k|k-1} H^T S_k^{-1} \quad (17)$$

використовується для побудови середнього фільтра та матриці:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (y_k - \hat{y}_k), \quad (18a)$$

$$P''_{k|k} = P_{k|k-1} - K_k S_k K_k^T. \quad (18b)$$

Результат $P''_{k|k}$ далі масштабується множником, що *нелінійно* залежить від y_k . Ступені свободи також зростають. Оновлення

$$P_{k|k} = \frac{\eta''_k + (y_k - \hat{y}_k)^T S_k^{-1} (y_k - \hat{y}_k)}{\eta''_k + m} P''_{k|k}, \quad (18c)$$

$$\eta_k = \eta''_k + m \quad (18d)$$

завершує рекурсію фільтра, даючи параметри

$$p(x_k | y_{1:k}) = \text{St}(x_k; \hat{x}_{k|k}, P_{k|k}, \eta_k). \quad (19)$$

Ступені свободи зростають після оновлення за вимірюванням, але редукуються до η'_k як перший крок часового оновлення. Якби не ця редукція, фільтр швидко збігся б до KF. Дійсно, множник у (18с) стає 1 при $\eta''_k \rightarrow \infty$, і залишається лише KF-оновлення. Отже, KF — одна з реалізацій наведеного фільтра. Це узгоджується зі збіжністю розподілу Стюдента t до гаусівського при нескінченних ступенях свободи.

Дві апроксимації (11) та (14) ведуть до t -щільностей у (12) та (15) і до фільтра, який нагадує КФ, окрім нелінійної залежності $P_{k|k}$ від y_k . Подібні апроксимації спільні для багатьох алгоритмів оцінювання стану.

Найпростіший вибір для користувача — припустити спільні ступені свободи для x_0, v_k, e_k з самого початку, тобто $\eta'_k = \eta_0 = \gamma = \delta$ дає $\eta''_k = \eta'_k$. Тоді єдине решта налаштування — перейти від η_k назад до η'_k після оновлення за вимірюванням.

5. Симуляційний приклад: супровід дрону (drone tracking)

t -фільтр і згладжувач тестуються на гіпотетичній задачі супроводу дрону. Розглянемо обмежену територію, що спостерігається кількома камерами — наприклад, двір промислового або урядового об'єкта. Огорожі та стіни можуть утримати пішого порушника, але дрон або безпілотний літальний апарат (БПЛА) важче перешкодити пройти. Фільтр супроводу на основі камер може використовуватись для, наприклад, ініціації тривоги. Порівняно зі супроводом великих літаків, дрони набагато маневреніші завдяки своєму розміру та керуванню. Вимірювання положення, отримані з детекцій кількох камер, можуть мати великі помилки через виклики обробки зображень (наприклад, рухомі дерева у фоні). Таким чином, приклад добре відповідає припущенням про важкохвостий шум.

Траєкторії маневрених дронів симулюються моделлю сталої швидкості [?] з чотиривимірним станом $x_k^T = [p_k^T, v_k^T]$, що складається з горизонтального положення та швидкості. Положення p_k вимірюється. Просторово-станова модель:

$$x_{k+1} = \begin{pmatrix} I_2 & T I_2 \\ 0 & I_2 \end{pmatrix} x_k + \begin{pmatrix} \frac{T^2}{2} I_2 \\ T I_2 \end{pmatrix} v_k, \quad (20a)$$

$$y_k = (I_2 \quad 0) x_k + e_k, \quad (20b)$$

де $T = 0,2$ с — крок дискретизації. Процесний шум $v_k = a_k$ — це нульове-середнє біле прискорення з гаусівським розподілом $\mathcal{N}(0, Q_k)$. Номінальна коваріація $Q_k = Q_{\text{nom}} = I_2/T^2$ виконується для більшості k . Маневри вводяться встановленням $Q_k = Q_{\text{man}} = 20^2 Q_{\text{nom}}$ для $k = 25, 75, 125$ (моменти часу $t = 5, 15, 25$ с). Вимірювання забруднюються викидами аналогічно. Вимірювальний шум $e_k = \tilde{p}_k$ — нульове-середнє гаусівський шум з $\mathcal{N}(0, R_k)$ і $R_k = R_{\text{nom}} = 5^2 I_2$ для більшості k , $R_k = R_{\text{out}} = 25^2$ для $k = 50, 100$ (моменти часу $t = 10, 20$ с). Це індукує великі помилки вимірювання у ці моменти.

Розглядається двір 300×300 м². Швидкість $s_k = \|v_k\|_2$ обмежена максимумом 30 м/с. Симулюються траєкторії з 151 кроку (30 с) з

початковим станом $x_0 = [150, 300, 0, -15]^T$ і приймаються лише ті, що задовольняють обмеження.

Порівнюються три фільтри:

1. KF зі знанням лише номінальних параметрів $Q_{\text{ном}}$ і $R_{\text{ном}}$.
2. “Ясновидний” KF (clairvoyant), що знає також $Q_{\text{ман}}$, R_{out} та моменти виникнення викидів. Це оптимальний фільтр, але він використовує знання, недоступне в реальних сценаріях.
3. t -фільтр (Sec. 4), що використовує лише номінальні параметри, але припускає t -шум з 3 ступенями свободи. Проміжні кроки апроксимації виконуються через мінімізацію KLD.

Продуктивність оцінюється помилкою положення $\|p_k - \hat{p}_k\|_2$. Типовий результат: для викиду, що супроводжується маневром:

- Ясновидний KF — не зазнає впливу ні від чого.
- Номінальний KF — має велику помилку положення, що повільно затухає.
- t -фільтр з номінальними параметрами — також зазнає великих помилок у моменти маневру та викиду, але його продуктивність відновлюється швидше за номінальний KF.

Аналогічний результат для згладжувачів: ясновидний RTS — найкращий, t -згладжувач — кращий за номінальний RTS у показовому прикладі.

Щоб підтвердити це для більшої кількості реалізацій, проведено 500 Монте-Карло симуляцій з обчисленням середньоквадратичної помилки

$$e = \left(\frac{1}{145} \sum_{k=5}^{150} \|p_k - \hat{p}_k\|_2^2 \right)^{1/2}. \quad (21)$$

Результат: t -фільтр покращує результат номінального KF.

Висновки експерименту.

- Гейтування (gating) у KF для відкидання викидів вимірювань часто веде до розходження, бо і маневри, і викиди дають великі залишки.
- Симуляції без маневрів і викидів дають подібну продуктивність для всіх алгоритмів — підтверджує властивість оптимальності.
- Вибір ступенів свободи у t -фільтрі має лише незначний вплив на помилку, якщо не обрано надто велике значення.

- Вибір масштабування матриць також вторинний для фільтра. Однак згладжувач без KLD-факторів масштабування погіршується.

Результати свідчать про потенціал фільтрування та згладжування Стюдента t як простих способів зробити KF та RTS-згладжувачі стійкішими.

6. Заключні зауваження

Ми дослідили використання розподілу Стюдента t як важкохвостової альтернативи гаусівському розподілу. Результати для розподілу Стюдента t та інших еліптично-контурних розподілів поставлено в контекст із поширеними виразами гаусівського розподілу. Точне обговорення фільтрування показало брак замкнених рекурсій для лінійних моделей з t -шумом.

Використовуючи дві проміжні t -апроксимації, виведено простий фільтр. Отримані вирази тісно нагадують фільтр Калмана, але містять нелінійне оновлення матричного параметра, що залежить від вимірювання. Крім того, обговорено розширення на нелінійні моделі та згладжувач. Симуляційні приклади свідчать про потенціал нашого підходу як простої, але стійкішої альтернативи фільтрові Калмана та згладжувачеві Рауха–Тунга–Штрібеля.