Introduction to Regularized DFA

Ginters Buss, *Riga Technical University*

***Abstract* – The paper studies regularization propeties of the regularized direct filter approach to high-dimensional filtering and real-time signal extraction. The regularized filter is endowed with three regularization terms for (i) coefficient smoothness, (ii) cross-sectional shrinkage and (iii) longitudinal shrinkage. Relative merits of the regularization terms are discussed. It is shown that the regularized filter is able to process high-dimensional data sets by controlling effective degrees of freedom using the three regularization terms, and that it is computationally fast.**

***Keywords* – high-dimensional filtering, real-time estimation, parameter shrinkage**

I. Introduction

Nowadays, the gathering of rich datasets is relatively easy. A more difficult exercise is to effectively use them for a particular problem at hand. This paper adds to the forecasting/ regularization/ shrinkage/ high-dimensional estimation literature (see e.g. ridge regression (e.g. Tikhonov and Arsenin, 1977; Hoerl and Kennard, 1970), lasso (Tibshirani, 1996), least angle regression (Efron et al, 2004), Bayesian shrinkage (e.g. Doan, Litterman and Sims, 1984), principal components (Stock and Watson, 2002)) by exploring the properties and abilities of a regularized direct filter approach (Wildi, 2012) in signal extraction and forecasting using many variables.

This paper is the first paper that studies and implements a regularized multivariate direct filter approach (Wildi, 2012). Filter regularization has found to help in real-time filter extraction since it controls for effective degrees of freedom; thus, it allows controling for overfitting that can have degrading effects in out-of-sample performance. Another advantage of a regularized filter is that it allows high-dimensional data to enter the filter and therefore further robustify the outcome. As it is shown in this paper, a particular regularization feature used in the paper may remind about the ‘lag decay’ term in Minnesota prior (see, e.g. Doan, Litterman and Sims, 1984) in Bayesian econometrics. Forcing more distant filter coefficients to zero both saves degrees of freedom and effectively shortens the filter, thus making it more responsive to changing environment. Another regularization feature studied in the paper is cross-sectional shrinkage that makes filter coefficients behave similarly for similar series. The cross-sectional shrinkage has been found to be useful particularly if the dataset is rather homogeneous.

II. Regularized Direct Filter Approach: An Overview

Denote y_T as the output of a symmetric, possibly bi-infinite filter, $\sum_{j=-\infty}^{\infty}\gamma_j L^j$, applied to input series x_T:

$$y_T = \sum_{j=-\infty}^{\infty}\gamma_j L^j x_T = \sum_{j=-\infty}^{\infty}\gamma_j x_{T-j}, \qquad (1)$$

where L is the lag or backshift operator. A real-time estimate of y_T is

$$\hat{y}_T = \sum_{j=0}^{T-1} b_j x_{T-j}. \qquad (2)$$

Denote the generally complex transfer functions of filters in (1) and (2) by $\Gamma(\omega) = \sum_{j=-\infty}^{\infty}\gamma_j \exp(-ij\omega)$ and $\hat{\Gamma}(\omega) = \sum_{j=0}^{T-1} b_j \exp(-ij\omega)$, respectively. For a stationary process x_T, the mean squared filter error (MSFE) can be expressed as the mean squared difference between the ideal output and the real-time estimate:

$$\int_{-\pi}^{\pi}|\Gamma(\omega)-\hat{\Gamma}(\omega)|^2 dH(\omega) = E[(y_T-\hat{y}_T)^2], \qquad (3)$$

where $H(\omega)$ is the unknown spectral distribution of x_T. A finite sample approximation of the MSFE, (3), is

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k |\Gamma(\omega_k)-\hat{\Gamma}(\omega_k)|^2 S(\omega_k), \qquad (4)$$

where $\omega_k = k2\pi/T$, $[T/2]$ is the greatest integer smaller or equal to $T/2$, and the weight w_k is defined as follows:

$$w_k = \begin{cases} 1 & \text{for}\,|k| \neq T/2 \\ 1/2 & \text{otherwise,} \end{cases} \qquad (5)$$

see Brockwell and Davis, 1987, Ch. 10 for the reason for w_k; although it is practically negligible, without it the inverse discrete Fourier transform does not replicate the data perfectly. This paper uses a ‘sufficient statistic’ – periodogram, $I_{Tx}(\omega_k)$ – as $S(\omega_k)$ in (4):

$$S(\omega_k) := I_{Tx}(\omega_k) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t \exp(-it\omega_k)\right|^2. \qquad (6)$$

Minimizing expression (4)-(6) yields the real-time filter output optimally approximated to the ideal output in mean squared error sense.

A. Univariate Direct Filter Approach

Rewrite discrete version MSFE, (4), as follows:

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k |\Gamma(\omega_k)-\hat{\Gamma}(\omega_k)|^2 I_{Tx}(\omega_k) W(\omega_k), \qquad (7)$$

which is identical to (4) for $W(\omega_k) := 1$. However, a more general version of $W(\omega_k) := W(\omega_k, expw, cut)$ can be written as follows:

$$W(\omega_k, expw, cut) = \begin{cases} 1 & \text{if}\,|\omega_k| < cut \\ (1+|\omega_k|-cut)^{expw} & \text{otherwise,} \end{cases} \qquad (8)$$

which collapses to unity for $expw = 0$, in which case the classical mean squared optimization, (4), is obtained. Parameter cut (for a ‘cut-off frequency’) marks the transition between the passband and rightmost stopband, and positive values of $expw$ (for ‘exponent weight’) emphasize high-frequency components in the rightmost stopband, thus making the filter output smoother than the one obtained by minimizing (4) for positive $expw$.

Univariate analysis is of limited usefulness; thus, we turn now to the multiple-series analysis.

B. Multivariate Direct Filter Approach

The above-mentioned univariate customized filter has been generalized to a multivariate filter in Wildi (2011). Rewrite univariate minimization problem, (7), with the discrete Fourier transform (DFT), $\Xi_{Tx}(\omega_k)$:

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k|\Gamma(\omega_k)-\hat{\Gamma}(\omega_k)|^2 I_{Tx}(\omega_k)W(\omega_k)$$

$$= \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k|\Gamma(\omega_k)\Xi_{Tx}(\omega_k)-\hat{\Gamma}(\omega_k)\Xi_{Tx}(\omega_k)|^2 W(\omega_k), \quad (9)$$

where

$$\Xi_{Tx}(\omega_k)=\sqrt{\frac{1}{2\pi T}}\sum_{t=1}^{T} x_t\exp(-it\omega_k). \quad (10)$$

In addition to the filter output, y_t, and the corresponding input, x_t, assume there are m additional explanatory variables z_{jt}, $j=1,\dots,m$ that may help improve the real-time estimate of y_t obtained with a univariate filter. Then, the second expression in the modulus on the second line of (9), $\hat{\Gamma}_X(\omega_k)\Xi_{Tx}(\omega_k)$, becomes

$$\hat{\Gamma}_X(\omega_k)\Xi_{Tx}(\omega_k)+\sum_{n=1}^{m}\hat{\Gamma}_{z_n}(\omega_k)\Xi_{Tz_n}(\omega_k), \quad (11)$$

where

$$\hat{\Gamma}_X(\omega_k)=\left(\sum_{j=0}^{L} b_{xj}\exp(-ij\omega_k)\right)\Xi_{Tx}(\omega_k) \quad (12)$$

$$\hat{\Gamma}_{z_n}(\omega_k)=\left(\sum_{j=0}^{L} b_{z_n j}\exp(-ij\omega_k)\right)\Xi_{Tz_n}(\omega_k) \quad (13)$$

are the one-sided transfer functions applied to the explanatory variables, and $\Xi_{Tx}(\omega_k)$, $\Xi_{Tz_n}(\omega_k)$ are the corresponding DFTs. Then, the multivariate version of (9) can be written as follows:

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k\left|\left(\Gamma(\omega_k)-\hat{\Gamma}_X(\omega_k)\right)\Xi_{Tx}(\omega_k)-\sum_{n=1}^{m}\hat{\Gamma}_{z_n}(\omega_k)\Xi_{Tz_n}(\omega_k)\right|^2 W(\omega_k). \quad (14)$$

C. Regularization

In order to conveniently define the regularized filter problem, the above-mentioned multivariate filtration problem is rewritten in a least squares form, see Wildi (2012) for details; this subsection explains how it is done, while the next subsection introduces the regularization problem.

Define X such that its k-th row, X_k, is:

$$X_{k'} = (1+I_{k>0})\mathrm{Vec}\begin{pmatrix}\Xi_{Tx}(\omega_k) & \cdots & \exp(-iL\omega_k)\Xi_{Tx}(\omega_k)\\ \Xi_{Tz_1}(\omega_k) & \cdots & \exp(-iL\omega_k)\Xi_{Tz_1}(\omega_k)\\ \Xi_{Tz_2}(\omega_k) & \cdots & \exp(-iL\omega_k)\Xi_{Tz_2}(\omega_k)\\ \vdots & \vdots & \vdots\\ \Xi_{Tz_m}(\omega_k) & \cdots & \exp(-iL\omega_k)\Xi_{Tz_m}(\omega_k)\end{pmatrix}, \quad (15)$$

where L is the filter length, and $I_{k>0}=0$ for $k=0$ and $I_{k>0}=1$ for $k=1,2,\dots,[T/2]$. Define vectors b and Y as follows:

$$b=\mathrm{Vec}\begin{pmatrix} b_{x0} & b_{z_10} & b_{z_20} & \cdots & b_{z_m0}\\ b_{x1} & b_{z_11} & b_{z_21} & \cdots & b_{z_m1}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ b_{xL} & b_{z_1L} & b_{z_2L} & \cdots & b_{z_mL}\end{pmatrix},$$

$$Y=\begin{pmatrix}\Gamma(\omega_0)\Xi_{Tx}(\omega_0)\\ 2\Gamma(\omega_1)\Xi_{Tx}(\omega_1)\\ 2\Gamma(\omega_2)\Xi_{Tx}(\omega_2)\\ \vdots\\ 2\Gamma(\omega_{[T/2]})\Xi_{Tx}(\omega_{[T/2]})\end{pmatrix}. \quad (16)$$

Neglecting a constant $2\pi/T$ and the practically negligible w_k, (14) with $W(\omega_k)=1$ can be rewritten as follows:

$$(Y-Xb)'(Y-Xb)\rightarrow\min_b \quad (17)$$

Since X and Y are complex-valued, the solution to (17) will also be complex-valued. A real-valued b can be obtained by rotating X and Y s.t. the value of the metric in (17) is unaffected:

$$X_{k,rot}=X_k\exp(-i\arg(\Gamma(\omega_k)\Xi_{Tx}(\omega_k))+ih\omega_k)$$
$$Y_{rot}=|Y|, \quad (18)$$

where $X_{k,rot}$ is the k-th row of X_{rot}, and h is the lag at which filter is estimated, i.e., $h=0$ for a concurrent filter that targets $y_{T-h}=y_T$, $h>0$ for a smoother, and $h<0$ for forecasting the signal. A real-valued b, thus, can be obtained from solving

$$(Y_{rot}-X_{rot}b)'(Y_{rot}-X_{rot}b)\rightarrow\min_b. \quad (19)$$

For the customized multivariate filter ($W(\omega_k)\neq 1$), define

$$X_{k,rot}^{cust}=X_{k,rot}\sqrt{W(\omega_k,expw,cut)} \quad (20)$$

$$Y_{rot}^{cust}=\begin{pmatrix}|\Gamma(\omega_0)\Xi_{Tx}(\omega_0)|\sqrt{W(\omega_0,expw,cut)}\\ 2|\Gamma(\omega_1)\Xi_{Tx}(\omega_1)|\sqrt{W(\omega_1,expw,cut)}\\ \vdots\\ 2|\Gamma(\omega_{[T/2]})\Xi_{Tx}(\omega_{[T/2]})|\sqrt{W(\omega_{[T/2]},expw,cut)}\end{pmatrix} \quad (21)$$

where $X_{k,rot}^{cust}$ is the k-th row of X_{rot}^{cust}. Then, the least-squares form for the customized filter problem can be written as follows:

$$(Y_{rot}^{cust}-X_{rot}^{cust}b)'(Y_{rot}^{cust}-X_{rot}^{cust}b)\rightarrow\min_b, \quad (22)$$

which collapses to (19) for $expw=0$.

Recalling that Tikhonov regularization problem (e.g. Tikhonov and Arsenin, 1977) can be cast in the form $(Y-Xb)'(Y-Xb)+\lambda b'b\rightarrow\min_b$, the regularized direct filter approach problem introduced in Wildi (2012) is of the familiar form:

$$(Y_{rot}^{cust}-X_{rot}^{cust}b)'(Y_{rot}^{cust}-X_{rot}^{cust}b)+\lambda_s b'Q_s b+\lambda_c b'Q_c b+\lambda_d b'Q_d b\rightarrow\min_b, \quad (23)$$

where the three additional expressions of bilinear form represent three different regularization directions – coefficient smoothness (subscript ‘s’), cross-sectional shrinkage (subsript ‘c’), and shrinkage along time dimension (subscript ‘d’). Let us discuss each in turn.

The idea behind the smoothness restriction is that filter coefficients should not change too erratically as functions of a lag. The Q_s matrix of size $(L+1)\times(L+1)$ is such that

$$b'Q_s b=\sum_{u=0}^{m}\sum_{l=2}^{L}\left((1-L)^2 b_l^u\right)^2, \quad (24)$$

where $(1-L)^2 b_l^u=b_l^u-2b_{l-1}^u+b_{l-2}^u$ is the second order difference of b_L^u, $l=0,\dots,L$, and $u=0,\dots,m$. Therefore, the term in (24) is a measure for the quadratic curvature of filter coefficients – if coefficients decay linearly as functions of a lag then this term vanishes. Thus, in the limiting case when

$\lambda_s \to \infty$, the filter coefficients are restricted to be linear functions of a lag.

The idea behind the cross-sectional shrinkage is that one will expect the filter coefficients to be similar for similar series. This shrinkage is implemented by imposing constraints on b according to

$$\sum_{u=0}^{m}\left(\left(b_0^u - \frac{1}{m+1}\sum_{u'=0}^{m} b_0^{u'}\right)^2 + \left(b_1^u - \frac{1}{m+1}\sum_{u'=0}^{m} b_1^{u'}\right)^2 + \dots + \left(b_L^u - \frac{1}{m+1}\sum_{u'=0}^{m} b_L^{u'}\right)^2\right) \tag{25}$$

which yields a symmetric bilinear form with

$$Q_c = \begin{pmatrix} q_{c,1} \\ q_{c,2} \\ \vdots \\ q_{c,(m+1)*(L+1)} \end{pmatrix} \tag{26}$$

where

$$\begin{aligned} q_{c,1} &= (1-\frac{1}{m+1},0,\dots,0| -\frac{1}{m+1},0,\dots,0| -\frac{1}{m+1},0,\dots,0|\dots) \\ q_{c,2} &= (0,1-\frac{1}{m+1},0,\dots,0\,|0,-\frac{1}{m+1},0,\dots,0| -\frac{1}{m+1},0,\dots,0|\dots) \\ q_{c,3} &= (0,0,-\frac{1}{m+1},0,\dots,0|0,0,-\frac{1}{m+1},0,\dots,0 \\ &|0,0,-\frac{1}{m+1},0,\dots,0|\dots) \\ &\dots \\ q_{c,(m+1)*(L+1)} &= (0,0,\dots,-\frac{1}{m+1}|0,0,\dots,-\frac{1}{m+1}|0,0,\dots, \\ &-\frac{1}{m+1}|\dots|0,0,\dots,1-\frac{1}{m+1}) \end{aligned} \tag{27}$$

such that each block separated by $|$ is of length $L+1$. Thus there are 1's on the diagonal of Q_c and periodically arranged $-\frac{1}{m+1}$'s which account for the central means in (25).

A higher λ_c gives preference for more similar filters across series and the limiting case, $\lambda_c \to \infty$ ensures the filter coefficients are identical across series.

Finally, the idea behind the shrinkage across time dimension is that a practitioner may give a preference for the filter coefficients that decay to zero progressively as functions of a lag. For a Bayesian econometrician this will remind of the lag decay in the Minnesota prior (e.g. Doan, Litterman and Sims, 1984). This shrinkage is implemented by setting Q_d such that

$$b'Q_d b = \sum_{u=0}^{m}\sum_{l=0}^{L}\tilde{q}_l (b_l^u)^2, \tag{28}$$

where $\tilde{q}_l$ is the l-th element of

$$\tilde{q} = \left(q^{0\vee h}, q^{|1-0\vee h|}, q^{|2-0\vee h|}, \dots, q^{|L-0\vee h|}\right), \tag{29}$$

where q is set to $q := 1+\lambda_d$, $\vee$ denotes a $\max(\cdot)$ function, and h signifies the lag at which filter is estimated, i.e., $h=0$ means a concurrent filter that targets $y_{T-h}=y_T$, $h>0$ means the filter is the smoother, and $h<0$ means the filter is targeted to forecast the signal h periods ahead. When estimating y_{T-h} for $h>0$ a practitioner will need to assign the largest filter weight to observations coinciding with y_{T-h}. Thus, (29) ensures that minimum regularization is imposed on lag h (since $q^{h-0\vee h}=q$), and a decay is emphasized symmetrically on both sides away from the target lag h. A higher λ_d ensures a faster coefficient decay to zero as a function of a lag.

Since the regularization is cast in bilinear forms, the problem in (23) has an analytic solution. Setting $\lambda_s=\lambda_c=\lambda_d=0$ gives the unregularized filter problem in (22). Or, setting $expw=0$ but letting some of the regularization lambdas be positive gives the regularized classical multivariate filter problem. This paper has found out that the lag decay shrinkage is the most useful of the three regularization types for the application at hand, followed by the cross-sectional shrinkage.

D. Filter Constraints

The first-order constraint imposes specific values for the amplitude functions at zero frequency. For a bandpass filter, one would typically set amplitudes at zero frequency to be zero ensuring that a bandpass filter attributes zero weight to the trend frequency, while for a univariate lowpass filter one would typically set the amplitude at zero frequency to unity to ensure that a lowpass filter tracks the level/scale of the target; such restriction is related to assuming the target has a unit root at zero frequency, i.e., it is the first-order integrated process.

For a multivariate filter, the optimal constrained level of the amplitude at zero frequency is less clear-cut. This level can be set to an inverse of the number of explanatory variables for all the variables, if all explanatory variables follow the same trend. However, the latter may not always be the case and thus a better outcome could be obtained by differentiating the amplitude constraint at zero frequency for various explanatory variables. An example of such differentiation of the constraint is provided in the empirical section.

In practice, one can choose to use or not to use the level constraint at one's own discretion. This constraint is implemented by restricting:

$$b_{-h}^u + b_{-(h-1)}^u + \dots + b_{L-h}^u = w^u, \tag{30}$$

where w^u is the value at which the transfer function for a variable u is set at zero frequency, and h is the targeted lag ($h=0$ for a concurrent filter, $h>0$ for a smoother, and $h<0$ for forecasting the signal).

The second-order constraint restricts the time shift of the filter at zero frequency to vanish, and is related to assuming the target variable has two unit roots at zero frequency, in which case both the first- and second-order constraints will be implemented. In practice, however, the usage of the constraints are up to the practitioner's agenda, and one could use the time shift constraint without imposing the level constraint, the combination of the constraints that can not be straightforwardly imposed in the time domain. The second-order constraint is imposed by forcing the derivative of the transfer function at zero frequency to vanish, which results in the following coefficient constraint:

$$-h\,b_{-h}^u + (1-h)\,b_{1-h}^u + (2-h)b_{2-h}^u + \dots + b_1^u + 2b_2^u + \dots + (L-h)b_{L-h}^u = 0, \tag{31}$$

where h is the targeted lag ($h=0$ for a concurrent filter, $h>0$ for a smoother, and $h<0$ for forecasting the signal).

Both constraints can be implemented by selecting any two of the coefficients but is implemented by constraining b_0^u and b_1^u, so as to avoid a conflicting situation between these constraints and the regularization, e.g., a lag decay agenda for h is large enough.

The constrained regularized filter problem is solved by rewriting filter coefficient vector b as follows:

$$b = Rb_f + c. \quad (32)$$

where b_f is the vector of freely determined filter coefficients, plugging (32) in (23), solving for b_f, and then plugging the estimate of b_f into (32) to get the estimate of b; see Wildi (2012) for details.

E. Effective Degrees of Freedom

In the unconstrained ordinary least squares framework, the (regression) degrees of freedom is the number of estimated parameters. Given a well-posed ordinary least squares problem,

$$(Y - Xb)'(Y - Xb) \rightarrow \min_b,$$

the fitted values of Y can be written in terms of a hat or smoother matrix, S, which is just a projection matrix, P:

$$\hat{Y} = SY = X(X'X)^{-1}X'Y = PY. \quad (33)$$

The degrees of freedom is trace of the projection matrix:

$$d.f. = tr(P), \quad (34)$$

which equals to $\mathrm{rank}(X)$.

For a regularized problem as in expression (23),

$$(Y_{rot}^{cust} - X_{rot}^{cust} b)'(Y_{rot}^{cust} - X_{rot}^{cust} b) + \lambda_s b'Q_s b + \lambda_c b'Q_c b$$

the smoother matrix is no longer an orthogonal projection but the same notion applies. Denoting the fitted value of Y_{rot}^{cust} by $\hat{Y}_{rot}^{cust}$ and the corresponding smoother matrix by $\hat{S}$:

$$\hat{S} = Re(X_{rot}^{cust})((X_{rot}^{cust})'X_{rot}^{cust} + \lambda_s Q_s + \lambda_c Q_c + \lambda_d Q_d)^{-1} Re(X_{rot}^{cust})', \quad (35)$$

such that $\hat{Y}_{rot}^{cust} = \hat{S}Y_{rot}^{cust}$, the effective degrees of freedom (or, effective number of parameters) is the trace of $\hat{S}$:

$$e.d.f. = tr(\hat{S}), \quad (36)$$

see, e.g. Moody (1992), Hodges and Sargent (2001).

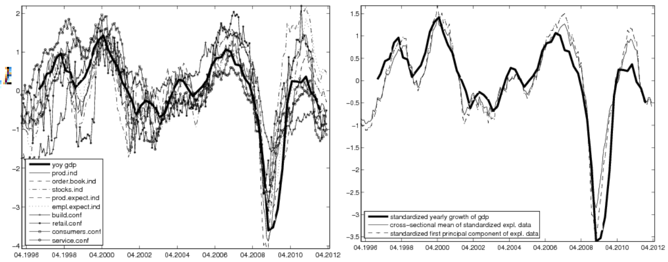
Effective degrees of freedom are useful for controlling the overfitting and, thus, for controlling out-of-sample performance.

III. Regularization Features

We now study the regularization features of the filter. For visual tractability and due to numerical issues (an unregularized filter crashes the high-dimensional input data when the number of estimated filter parameters reaches the number of sample observations) only nine variables are used to analyse the filter effect. The nine variables are business and consumers confidence data: production trend observed in recent months (industry), assessment of order-book levels (industry), assessment of stocks of finished products (industry), production expectations for the months ahead (industry), employment expectations for the months ahead (industry), confidence indicator in construction, confidence indicator in retail, consumer confidence indicator, and confidence indicator in services. These variables are used to track a lowpass in the euro-area gross domestic product (GDP).

In order to motivate the chosen transformation of data, it is illustrative to plot the transformed target variable and explanatory variables. Fig. 1a shows standardized annual growth of EA GDP versus standardized business and consumer data. Explanatory data are well aligned with the annual growth of GDP. Extracting the cross-sectional mean and the first principal component of the standardized explanatory data and plotting against standardized annual growth of GDP shows that both the mean and the first principal component explain annual changes in GDP well, and there is not much difference in the performance of the mean versus the principal component, see Fig. 1b.

Clearly, there is not much to improve upon the simple cross-sectional mean or the first principal component of the explanatory variables as it comes to tracking cyclical developments in the normalized annual growth of euro-area GDP; it is slightly more difficult to track non-normalized target, see the results below. The cross-sectional mean or principal components can be used as filter inputs, but this paper shows that it is not necessary to do so and that one can use the original, possibly high-dimensional data as the input and potentially benefit from the richness of data.



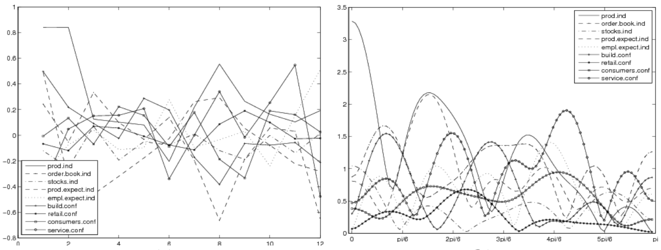
(a) (b)

Fig. 1: (a) Annual growth of GDP versus business and consumer data, all normalized to zero mean and unit variance. (b) Annual growth of GDP versus the cross-sectional mean and the first principal component of business and consumer data, all normalized to zero mean and unit variance.

In order to understand the extent of overparameterization in an unregularized multivariate filter, consider an unconstrained filter applied to the above-mentioned nine variables targeting an ideal lowpass of annual growth of EA GDP with the cut-off wave length of 12 months. The filter length is set to be fixed 12 observations, for simplicity. While the estimation routine can estimate a 9-variable filter on the full sample (178-observation long), it crashes smaller subsamples because of the degrees of freedom having been shrunk to zero for all subsamples shorter than 9*12=108 observations. A further reduction of filter length may be a temporary solution but not for long and not without consequences on the output quality. Therefore, an unconstrained 9-variable filter output is infeasible for the considered data samples. Thus, some sort of parameter shrinkage is necessary. In order to illustrate the effect of the parameter shrinkage induced by the regularized filter, consider the estimated filter coefficients for an unconstrained and unregularized 9-variable filter on the full sample. The number of estimated parameters is 9 variables and 12 observations, which gives 108 parameters to estimate on a 178-observation long sample, which gives only 70 residual degrees of freedom. Fig. 2a shows that the estimated filter coefficients look erratic, unsmooth and do not show either a similar behaviour between the variables or an evident decay towards zero with an increasing lag. Fig. 2b shows the (rather chaotic) filter amplitudes corresponding to the coefficients in Fig. 2a; it will be useful to analyse how the amplitudes change with various constraints and regularization restrictions.

We will now witness the effect of filter constraints and the regularization features first applied to each of them at a time and then in a potentially useful combination.

The first-order restriction imposes the filter amplitude to be a specific value at zero frequency. For a univariate lowpass filter, a natural value of the amplitude at zero frequency is unity in order to ensure that the scale of the output is comparable to the scale of the target signal. For a multivariate filter, the situation is



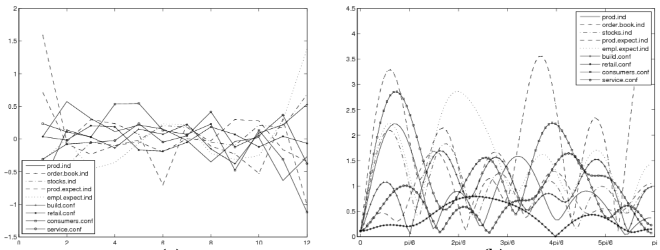
(a) (b)

Fig. 2: (a) 9-variable filter coefficients without regularization and filter constraints. The estimated filter coefficients look erratic, unsmooth and do not show a similar behaviour between the variables nor an evident decay towards zero with an increasing lag coefficients. (b) Filter amplitudes corresponding to the coefficients in Fig. 2a.

not so straightforward since all the input series generally do not possess the same trend, therefore restricting all amplitudes to be of the same value at zero frequency may be suboptimal. If all the input series followed a common trend, then it would be natural for a multivariate lowpass to set amplitudes at zero frequency to be inverse of the number of input series, so that summing over the amplitudes would result in unity at zero frequency. Since the input series used in this research have a somewhat similar behaviour between each other, the latter approach is used in this case; however, there may be potential gains by using a more sophisticated amplitude constraint that will differentiate amplitude values at zero frequency for different input series; such an approach is discussed later in the section when applying the filter to a higher dimensional set of explanatory variables.

The first-order constraint saves one d.f. per input series; thus, nine d.f. are saved for an unregularized nine-variable filter.

Fig. 3a and 3b show that the effect of amplitude constraint results in slightly more dispersed coefficients (the scale of the graph has changed), as well as slightly more exploded amplitudes. Thus, the first-order constraint per se does not seem to be of much help for an ill-posed high-dimensional filter. Note that the amplitude constraint is binding for almost all series since the unconstrained amplitudes at zero frequency are dispersed far away from the constrained value (1/9).



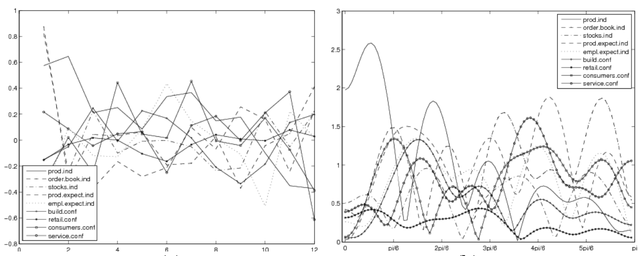
(a) (b)

Fig. 3: (a) Coefficients for a first-order constrained lowpass filter. (b) Filter amplitudes corresponding to the coefficients in Fig. 3a.

The second-order restriction imposes a vanishing phase shift at zero frequency for a targeted lead or lag, and also saves a d.f. per input variable in an unregularized problem. This constraint is related to assuming the target variable follows the second-order integrated process, in which case there are two unit roots at zero frequency and, therefore, both first- and second-order constraints will be implemented. However, the time-shift constraint can be used without the first-order constraint in order to ensure the output is coincident with the target signal but not necessarily assuming that the target signal follows a second-order integrated process. Therefore, such a combination of constraints goes beyond the one typically seen in the time-domain applications.

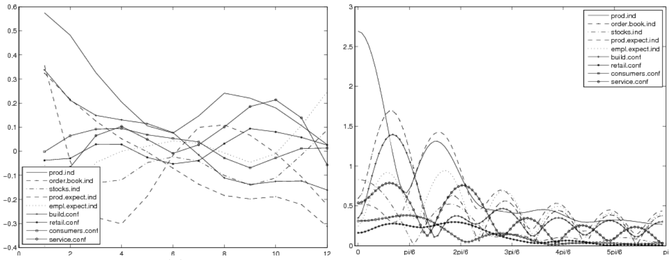
The corresponding filter coefficient and amplitude (see Fig. 4a and 4b) show that the coefficients are back to their original scale and also amplitudes look less exploded compared to the ones of the first-order constrained filter. (Evidently, higher amplitudes at the high-frequency content indicate that zero time shift at zero frequency is obtained by attributing higher weight to the high-frequency content, which is typically the case, when the explanatory variables are lagging with respect to the target variable, which is in line with the observation from Fig. 1a and 1b.) Still, the second-order constraint is not a panacea since the amplitudes are still erratic and since the number of degrees of freedom vanishes for samples smaller than 9*(12-1)=99 months, which is 8 years of data.

Turning to the new regularization features, Fig. 5 to 7 show the effect of coefficient smoothness restriction of various extent corresponding to λ_s being 0.01, 0.1 and 1, which correspond to the effective degrees of freedom 66, 43 and 30, respectively.



(a) (b)

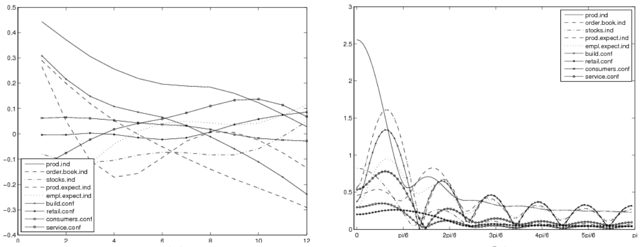
Fig. 4: (a) Coefficients for a second-order constrained concurrent filter. (b) Filter amplitudes corresponding to the coefficients in Fig. 4a.



(a)

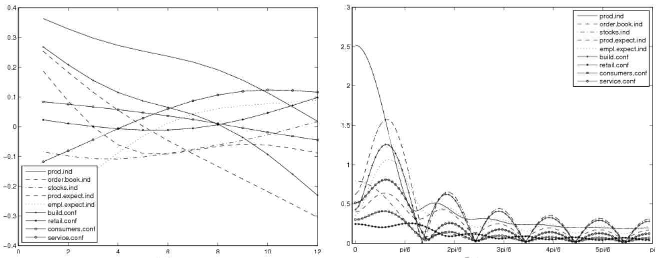
(b)

Fig. 5: (a) Coefficients for an unconstrained filter if $\lambda_s = 0.01$. (b) Filter amplitudes corresponding to the coefficients in Fig. 5a.



(a) (b)

Fig. 6: (a) Coefficients for an unconstrained filter if $\lambda_s = 0.1$. (b) Filter amplitudes corresponding to the coefficients in Fig. 6a.

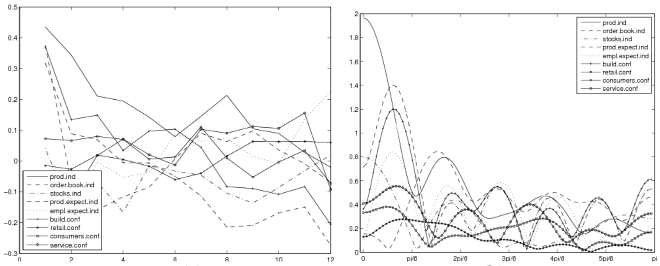


(a) (b)

Fig. 7: (a) Coefficients for an unconstrained filter if $\lambda_s = 1$. (b) Filter amplitudes corresponding to the coefficients in Fig. 7a.

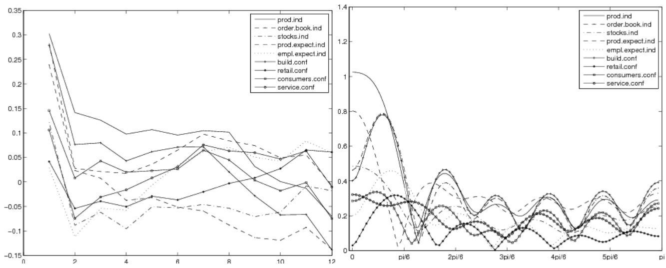
Fig. 5 to 7 show that the filter coefficients are no longer erratic; they are nice and smooth and they are getting more linear as the smoothness parameter λ_s increases. If the smoothness parameter is increased still further, the filter coefficients converge to horizontal straight lines. However, such an over-regularization is not necessary or desirable since the considered small values of the smoothness tuning coefficient already reduce a lot of degrees of freedom and the corresponding amplitudes look much closer to those that are expected, i.e., most of their weights concentrate on the passband $[0, \pi/6]$ and converge to zero in the stopband. Nonetheless, the filter coefficients show neither convergence to zero with higher lags, nor similarity across series.

Fig. 8 to 10 show the (partial) effect of cross-sectional restriction of various extent corresponding to λ_c being 0.01, 0.1 and 1 (the rest of shrinkage parameters being zero), which correspond to the effective degrees of freedom 85, 48 and 24, respectively, which is close to that observed with parameter smoothness restriction.



(a) (b)

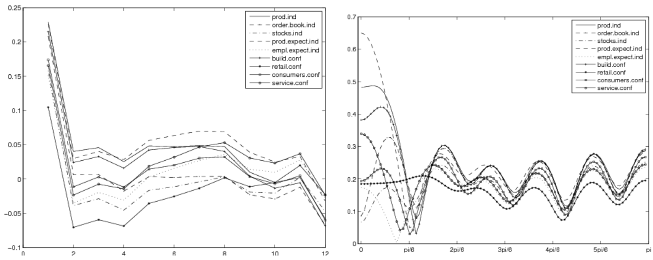
Fig. 8: (a) Coefficients for an unconstrained filter if $\lambda_c = 0.01$. (b) Filter amplitudes corresponding to the coefficients in Fig. 8a.



(a) (b)

Fig. 9: (a) Coefficients for an unconstrained filter if $\lambda_c = 0.1$. (b) Filter amplitudes corresponding to the coefficients in Fig. 9a.

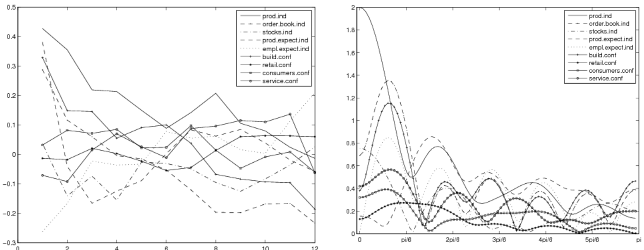
The effects of cross-sectional restriction differ from those of parameter smoothness restriction – mild cross-sectional restriction seemingly improves the behaviour of filter coefficients and amplitudes (see Fig. 8a and 8b), but further cross-sectional restriction can be harmful if applied alone (see amplitude behaviour in Fig. 10b). Such a cross-sectional restriction analysis may help understand which series or clasters of series are different from the others. In our research, no series clearly stands out from the rest.



(a) (b)

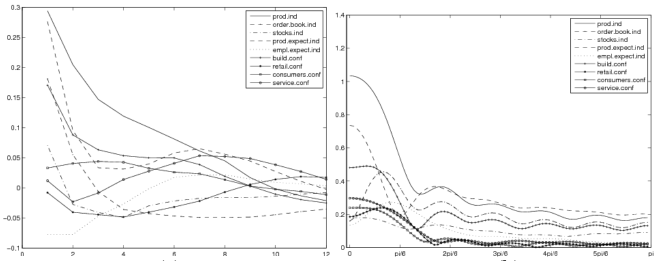
Fig. 10: (a) Coefficients for an unconstrained filter if $\lambda_c = 1$. (b) Filter amplitudes corresponding to the coefficients in Fig. 10a.

As for the third regularization feature, Fig. 11 to 13 show the longitudinal effect, i.e. a lag decay restriction of various extent corresponding to λ_d being 0.01, 0.1 and 1, which correspond to the effective degrees of freedom 82, 30 and 5, respectively, which is a stronger shrinkage than that observed with parameter smoothness or cross-sectional restriction.



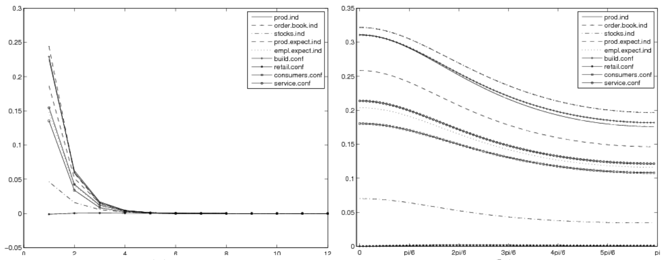
(a) (b)

Fig. 11: (a) Coefficients for an unconstrained filter if $\lambda_d = 0.01$. (b) Filter amplitudes corresponding to the coefficients in Fig. 11a.



(a) (b)

Fig. 12: (a) Coefficients for an unconstrained filter if $\lambda_d = 0.1$. (b) Filter amplitudes corresponding to the coefficients in Fig. 12a.



(a) (b)

Fig. 13: (a) Coefficients for an unconstrained filter if $\lambda_d = 1$. (b) Filter amplitudes corresponding to the coefficients in Fig. 13a.

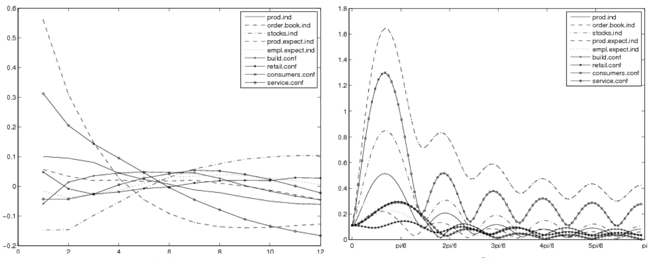
Fig. 11 to 13 show that a lag decay restriction forces filter coefficients to shrink towards zero as functions of lag and that a sufficiently high shrinkage parameter yields filter coefficients to be non-zero for a small number of lags. Fig. 13 shows that a sufficiently high longitudinal shrinkage forces filter amplitudes to shrink towards zero (see the scale of Fig. 13b) and flatten, resembling those of an allpass filter, which is an expected behaviour since a short filter cannot discriminate between frequencies effectively.

Coefficients in Fig. 11a and 12a are rather smooth, which resembles the effect of parameter smoothness restriction. Also, Fig. 11a and 12a show that longitudinal restriction forces filter coefficients to behave somewhat similarly across series, which reminds of the cross-sectional shrinkage. These effects might suggest that the lag decay shrinkage is the most useful of all three shrinkages. Still, the longitudinal shrinkage may conflict with e.g. parameter smoothness restriction for a sufficiently high lag decay restriction, see Fig. 13a. However, instead of using both longitudinal and parameter smoothness regularization features, one may just loosen the lag decay restriction.

The findings of this paper, indeed, suggest that the longitudinal shrinkage may be the most useful of the three regularization features. Moreover, this paper will use only the longitudinal and the cross-sectional shrinkages from the considered regularization 'troika' since the parameter smoothness restriction can be obtained implicitly by the former two.

Recall that setting the longitudinal shrinkage to $\lambda_d = 1$ yields only five e.d.f., which may suggest that a slight change in the sample size or in the number of explanatory series can yield close to zero e.d.f. Indeed, the estimation routine can break up if severe regularization is imposed. Therefore, a caution should be taken in empirical work so that a sufficient number of effective degrees of freedom are given to the estimation routine. Otherwise, the estimation routine will not work not because of overparameterization but because of 'underparameterization'.

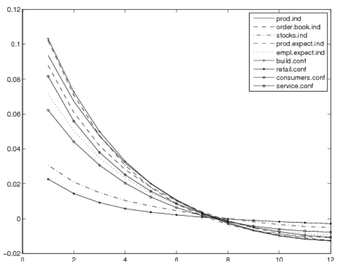
Filter constraints have been found to be useful in real-time signal extraction (see e.g. Buss, 2012). Therefore, consider the effect of longitudinal shrinkage combined with the first-order constraint or second-order constraint or both first- and second-order constraints.

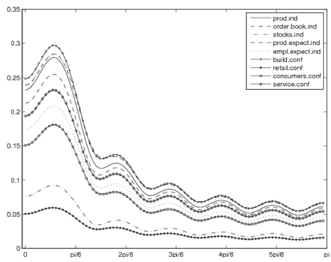


(a) (b)

Fig. 14: (a) Coefficients if longitudinal regularization with $\lambda_d = 0.1$ and the first order constraint are implemented. (b) Filter amplitudes corresponding to the coefficients in Fig. 14a.

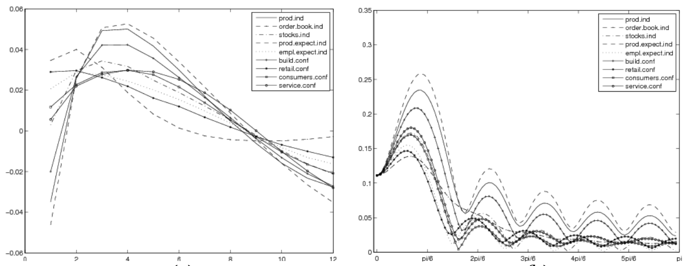
Implementing the first-order constraint together with the longitudinal shrinkage yields similarly-behaved coefficients and amplitudes, whose values at zero frequency are an inverse of the number of input variables, i.e 1/9. Amplitude values tend to diverge sharply and mostly increase for passband frequencies, after which they tend to converge and decrease.





(a) (b)

Fig. 15: (a) Coefficients if longitudinal regularization with $\lambda_d = 0.1$ and the second order constraint are implemented. (b) Filter amplitudes corresponding to the coefficients in Fig. 15a.



(a) (b)

Fig. 16: (a) Coefficients if longitudinal regularization with $\lambda_{2} = 0.1$ and both the first- and the second-order constraints are implemented. (b) Filter amplitudes corresponding to the coefficients in Fig. 16a.

The instability of the amplitudes at low frequencies may be explained by the restrictive nature of the first-order constraint – it forces all amplitudes to be of the same small value although the unrestricted amplitudes are somewhat dispersed around zero frequency. Also, some of the coefficients are negative at low lags, which can be considered an undesirable effect for the dataset, where each series correlates positively with the target.

The second-order constraint slightly increases the dispersion of the coefficients but otherwise does not add drastic changes to the regularized filter.

Implementing both constraints simultaneously is the most restrictive case. Fig. 16a and 16b show that filter coefficients behave more similarly among series than in the case of no constraints or just the first-order constraint (notice the scale of graphs), and so the corresponding amplitudes are less dispersed than in the case of no constraints or just the first order constraint. Still, negative coefficient values implied by the first-order constraint may be considered somewhat implausible/undesirable, as well as the cause of their implausibility – the restrictive and somewhat arbitrary amplitude constraint. Therefore, if the first-order constraint is to be used, one should think of plausible values for amplitudes at zero frequency. Otherwise, the practitioner may be willing to use the cross-sectional shrinkage as a tool to help controlling the degrees of freedom (at least for rather homogeneous datasets), instead of using the amplitude constraint.

IV. Conclusions

Nowadays, information is abundant. Statistical tools are being developed that are suitable to process a large amount of information for a particular problem at hand. This paper has considered the regularized multivariate direct filter approach (Wildi, 2012) as a tool for signal extraction and forecasting using high-dimensional datasets. The paper has studied the regularization properties of the filter: (i) coefficient smoothness, (ii) cross-sectional shrinkage, and (iii) longitudinal shrinkage. Relative merits of the three regularization terms have been discussed. It has been shown that the filter can be successfully applied to high-dimensional datasets.

References

[1] Brockwell, P. J. and R. A. Davis (1987), *Time Series: Theory and Methods*, New York: Springer Verlag.

[2] Buss, G. (2012), "A new real-time indicator for the Euro Area GDP", Working Papers 2012/02, Latvijas Banka.

[3] Doan, T., R. B. Litterman and C. A. Sims (1984), "Forecasting and conditional projection using realistic prior distributions", *Econometric Reviews*, vol. 3(1), 1-100.

[4] Efron, B., T. Hastie, I. Johnstone and R. Tibshirani (2004), "Least angle regression", *The Annals of Statistics*, vol. 32(2), 407-499.

[5] Hodges, J. S. and D. J. Sargent (2001), "Counting degrees of freedom in hierarchical and other richly-parameterized models", *Biometrika*, vol. 88(2), 367-379.

[6] Hoerl, A. E. and R. W. Kennard (1970), "Ridge regression: applications to nonorthogonal problems", *Technometrics*, vol. 12(1), 69-82.

[7] Moody, J. E. (1992), "The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems", *Advances in Neural Information Processing Systems 4* (eds J. E. Moody, S. J. Hanson and R. P. Lippmann), 847-854, San Mateo: Morgan Kaufmann.

[8] Stock, J. H. and M. W. Watson (2002), "Macroeconomic forecasting using diffusion indexes", *Journal of Business & Economic Statistics*, vol. 20(2), 147-162.

[9] Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 58(1), 267-288.

[10] Tikhonov, A. N. and V. Y. Arsenin (1977), *Solutions of Ill-Posed Problems*, Washington: V. H. Winston & Sons.

[11] Wildi, M. (2008), *Real-Time Signal-Extraction: Beyond Maximum Likelihood Principles*.

[12] Wildi, M. (2009), ``Real-Time US-Recession Indicator (USRI): A Classical Cycle Perspective with Bounceback", IDP-Working Paper, IDP-WP-09Jun-06.

[13] Wildi, M. (2011), "I-DFA and I-MDFA: Companion paper to R-code published on SEFBlog", IDP-Working Paper.

[14] Wildi, M. (2012), "Elements of forecasting and signal extraction", IDP-Working Paper.

Ginters Buss was born in Latvia. He received the Bachelor Degree in Economics from the University of Latvia, studied at Ghent University, Belgium and later at the Central European University, Hungary, where he obtained his Master Degree in Economics in 2007. He was a scholar at Duke University, USA for a year; currently he is a doctoral student at Riga Technical University. He is a Member of the Latvian Mathematical Society.
E-mail: ginters.buss@gmail.com

Ginters Bušs. Ievads regularizētā tiešā filtra pieejā

Šis darbs pēta regularizēta tiešā filtra pieeju (Vildi, 2012) kā rīku lieldimensionālu datu filtrēšanai un reālā laika signālu iegūšanai. Lai gan Vildi (2012) aprēķina filtru, tas neanalizē filtra regularizācijas īpašības uz reāliem datiem, kā arī nepēta iespējas pielietot filtru lieldimensionālām datu kopām. Tādējādi šā pētījuma jauninājums ir padziļināta filtra regularizācijas īpašību izpēte, kas ļauj pielietot filtru uz reāliem, potenciāli lieldimensionāliem datu masīviem. Vildi (2012) regularizē filtru, kas izstrādāts Vildi (2011). Filtra regularizācija ietver Vildi (2011) neregularizētā filtra optimizācijas problēmas pārveidošanu mazāko kvadrātu izteiksmē un trīs regularizācijas locekļu ieviešanu, kas atbild par filtra (i) koeficientu gludumu, (ii) šķērsgriezuma sašaurināšanu un (iii) garengriezuma sašaurināšanu. Koeficientu gluduma regularizācija nodrošina, ka filtra koeficienti mainās gludi laika dimensijā. Pastiprināta gluduma regularizācija nodrošina, ka koeficienti kļūst arvien lineārāki laika griezumā. Šķērsgriezuma regularizācija nodrošina filtra koeficientu līdzīgu uzvedību starp līdzīgiem ievaddatiem. Robežgadījumā koeficienti ir vienādi visiem ievaddatiem. Garengriezuma sašaurināšana nodrošina filtra koeficientu dilšanu līdz nullei līdz ar augstāku lagu. Rezultāti rāda, ka visefektīvākais regularizācijas loceklis ir garengriezuma sašaurināšana, kam seko šķērsgriezuma sašaurināšana. Secināts, ka regularizētais filtrs ir spējīgs apstrādāt lieldimensionālas datu kopas, kontrolējot efektīvās brīvības pakāpes, un ka filtra skaitļošana ir ātra. Tāpat secināts, ka regularizēta tiešā filtra pieeja ir vērtīga gan vienlaicīgai novērtēšanai, gan prognozēšanai, izmantojot lieldimensionālus datus.

Гинтер Буш. Введение в подход регуляризованного прямого фильтра

Эта работа исследует подход регуляризованного прямого фильтра (Вилди, 2012) в качестве инструмента фильтрации многомерных данных и получения сигналов в режиме реального времени. Хотя в Вилди (2012) рассчитан фильтр, эта работа не анализирует свойства регуляризации на реальных данных, а также не исследует возможности использования фильтра для многомерных данных. Таким образом, новизна данного исследования состоит в углубленном изучении свойств регуляризации фильтра, которые позволяют применять фильтр к реальным, потенциально многомерным данным. Вилди (2012) регуляризует фильтр, созданный в Вилди (2011). Регуляризация включает трансформацию задачей оптимизации нерегуляризованного фильтра Вилди (2011) в условия наименьших квадратов и введении трех членов регуляризации, которые отвечают за (i) гладкость коэффициентов, (ii) поперечную усадку и (iii) продольную усадку. Гладкость коэффициентов гарантирует, что коэффициенты фильтра плавно изменяются во временной дименсии. Повышенная гладкость регуляризации обеспечивает все большую линейность коэффициентов во временной дименсии. Поперечная усадка гарантирует схожее поведение коэффициентов фильтра при схожих входных данных. В предельном случае коэффициенты одинаковы для всех данных. Продольная усадка гарантирует, что коэффициенты фильтра стремятся к нулю при более далеких отставаниях. Результаты показывают, что наиболее эффективным методом регуляризации является продольная усадка, вторым по эффективности – поперечная усадка. Сделан вывод, что регуляризованный фильтр способен обрабатывать наборы многомерних данных, используя управление эффективными степенями свободы, и что вычисления фильтрапроводятся оперативно. Также установлено, что подход регуляризованного прямого фильтра является ценным как для одновременной оценки, так и для прогнозирования при использовании многомерных данных.