

Image Restoration Using Noisy ICA, PCA Compression and Code Shrinkage Technique

Conf.dr. Cătălina-Lucia COCIANU
Catedra de Informatică Economică, A.S.E. București

The research reported in the paper aims the development of some methodologies for noise removal in image restoration. In real life, there is always some kind of noise present in the observed data. Therefore, it has been proposed that the ICA model used in image restoration should include noise term as well. Different methods for estimating the ICA model when noise is present have been developed. In noisy ICA, we have to deal with the problem of estimation of the noise free realization of the independent components. The noisy ICA model can be used to develop a denoising method, namely the sparse code shrinkage [10]. The final part of the paper presents a LMS optimal PCA compression/decompression scheme, where the noise is annihilated in the feature space. In order to derive conclusions concerning the correlations between the dimensionality reduction and the resulted quality of the restored images as well as the effect of using both LMS optimal compression/decompression technique and PCA based noise removal method several tests were performed on the same set of data. The tests proved that the proposed restoration technique yields high quality restored images in both cases, when the CSPCA algorithm was applied directly on the initial image and when it was applied in the reduced feature space respectively. [21]

Keywords: ICA, noisy ICA, feature extraction, PCA, image processing, data restoration, noise removal, shrinkage function

1 Introducere

În multe aplicații practice, datele observate conțin o componentă de tip zgomot, care poate fi rezultatul modului de observare /achiziționare a datelor sau al alegerii inadecvate a modelului. În această situație, sunt utilizate modelele ICA cu zgomot înglobat aditiv.

Definiția unui model ICA cu zgomot este,

$$(1) \mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\eta},$$

unde $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$ este vectorul zgomot. În general, tehnicile ICA în cadrul modelelor cu zgomot sunt dezvoltate în ipotezele în care zgomotul este gaussian și independent de componentele principale $s_i, 1 \leq i \leq n$. Matricea de covarianță a zgomotului este în general considerată cunoscută sau de forma $\boldsymbol{\Sigma}_\eta = \sigma^2 \mathbf{I}$. Realizările componentelor independente $s_i, 1 \leq i \leq n$ nu mai pot fi identificate, deoarece nu pot fi complet separate de zgomot [9].

În cazul în care matricea de covarianță a vec-

torului zgomot este $\boldsymbol{\Sigma}_\eta = \sigma^2 \mathbf{I}$, vectorul zgomot din (1) este considerat de tip senzor: variabilele zgomot sunt adăugate separat fiecărei variabile observate $x_i, 1 \leq i \leq n$.

Dacă vectorul zgomotul este adăugat, componentă cu componentă, variabilelor $s_i, 1 \leq i \leq n$, atunci zgomotul este de tip sursă. Modelele ICA în care zgomotul este de tip sursă sunt descrise prin intermediul relației,

$$(2) \mathbf{x} = \mathbf{A}(\mathbf{s} + \boldsymbol{\eta}),$$

unde covarianța vectorului zgomot este considerată matrice diagonală.

În situația modelelor ICA cu zgomot sursă, componentele independente sunt,

$$\tilde{s}_i = s_i + \eta_i, 1 \leq i \leq n,$$

și modelul ICA (2) este,

$$(3) \mathbf{x} = \mathbf{A}\tilde{\mathbf{s}}.$$

Relația (3) definește un model ICA standard în situația în care componentele $\tilde{s}_i, 1 \leq i \leq n$ sunt independente și non-gaussiene, deci modelul poate fi estimat utilizând tehnicile

corespunzătoare modelelor ICA fără zgomot.

În ipoteza în care matricea de covarianță a vectorului zgomot este $\Sigma_\eta = \mathbf{A}\mathbf{A}^T\sigma^2$, principiul care stă la baza estimării modelului (1) este următorul. Fie $\tilde{\boldsymbol{\eta}} = \mathbf{A}^{-1}\boldsymbol{\eta}$; obținem,

$$\mathbf{x} = \mathbf{A}(\mathbf{s} + \mathbf{A}^{-1}\boldsymbol{\eta}) = \mathbf{A}(\mathbf{s} + \tilde{\boldsymbol{\eta}}).$$

Covarianța vectorului zgomot $\tilde{\boldsymbol{\eta}}$ este matricea $\Sigma_{\tilde{\eta}} = \sigma^2\mathbf{I}$, deci componentele independente transformate sunt independente. Rezultă că în această situație matricea \mathbf{A} poate fi estimată prin utilizarea tehnicilor corespu-zătoare modelelor ICA fără zgomot.

2. Estimarea modelului ICA cu zgomot.

Tehnici de eliminare a interferențelor

Metodele de eliminare a interferențelor presupun modificarea tehnicilor de estimare a modelelor ICA standard astfel încât zgomotul să fie eliminat sau cel puțin redus. Fie modelul ICA standard,

$$(4) \mathbf{v} = \mathbf{A}\mathbf{s}.$$

Componentele independente ale modelului (4) sunt determinate prin proiecțiile $\mathbf{w}^T\mathbf{v}$, \mathbf{v} vector alb și astfel încât non-gaussianitatea este maximizată local în ipoteza $\|\mathbf{w}\|=1$. Obținem, $\mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{v} + \mathbf{w}^T\boldsymbol{\eta}$, deci ideea este de a măsura non-gaussianitatea proiecției $\mathbf{w}^T\mathbf{v}$ pe baza proiecției observate $\mathbf{w}^T\mathbf{x}$ astfel încât măsura să nu fie afectată de zgomotul $\mathbf{w}^T\boldsymbol{\eta}$.

Eliminarea interferențelor prin kurtosis

Deoarece măsura kurtosis este imună la zgomotul gaussian, $\text{kurt}(\mathbf{w}^T\mathbf{x}) = \text{kurt}(\mathbf{w}^T\mathbf{v})$, poate fi aplicată în cadrul modelelor ICA cu zgomot.

Algoritmii de tip gradient pentru maximizarea măsurii kurtosis sunt de obicei aplicați vectorilor albi. Pentru estimarea componentelor independente ale modelelor ICA cu zgomot este luat în calcul și efectul transformării vectorului datelor observate \mathbf{x} în vector alb asupra componentei zgomot. Dacă matricea de covarianță a vectorului zgomot Σ_η este cunoscută și $\Sigma = \mathbf{E}(\mathbf{x}\mathbf{x}^T)$ (\mathbf{x} este centrat), atunci operația de transformare a lui \mathbf{x} în vector alb este înlocuită cu,

$$(5) \tilde{\mathbf{x}} = (\Sigma - \Sigma_\eta)^{-\frac{1}{2}}\mathbf{x},$$

cu alte cuvinte, în loc de Σ este utilizată matricea de covarianță a datelor fără zgomot, $\Sigma - \Sigma_\eta$.

După efectuarea operației (5), datele observate sunt modelate prin, $\tilde{\mathbf{x}} = \mathbf{B}\mathbf{s} + \tilde{\boldsymbol{\eta}}$, unde \mathbf{B} este ortogonală și $\tilde{\boldsymbol{\eta}}$ este transformata liniară a vectorului zgomot $\boldsymbol{\eta}$.

Eliminarea interferențelor utilizând măsurile generale ale non-gaussianității

În continuare vom considera măsura,

$$(6) J_G(\mathbf{w}^T\mathbf{v}) = [E(G(\mathbf{w}^T\mathbf{v})) - E(G(v))]^2,$$

unde G este o funcție nepolinomială și v este variabilă aleatoare gaussiană standard. Măsura definită prin (6) poate fi utilizată pentru estimarea modelelor ICA cu zgomot dacă $J_G(\mathbf{w}^T\mathbf{v})$ poate fi estimată pe baza observațiilor perturbate \mathbf{x} . Cu alte cuvinte, dacă z este o variabilă aleatoare non-gaussiană și η este o variabilă zgomot distribuită normal, cu medie 0 și varianță σ^2 , atunci relația dintre $E(G(z))$ și $E(G(z+\eta))$ trebuie să poată fi exprimată cât mai simplu.

Fie

$$\varphi_c(x) = \frac{1}{c} \varphi\left(\frac{x}{c}\right) = \frac{1}{\sqrt{2\pi}c} \exp\left(-\frac{x^2}{2c^2}\right)$$

densitatea de repartiție gaussiană de medie 0 și varianță c^2 și, pentru orice număr natural k , fie $\varphi_c^{(k)}(x)$ derivata de ordin k a funcției

$$\varphi_c(x), \text{ respectiv } \varphi_c^{(-k)}(x) = \int_0^x \varphi_c^{(-k+1)}(y)dy,$$

unde $\varphi_c^{(0)}(x) = \varphi_c(x)$.

Estimarea componentelor independente pe baza observațiilor cu zgomot poate fi realizată prin maximizarea funcției definite de (6) [9], [10]; estimarea statisticii $E(G(\mathbf{w}^T\mathbf{v}))$ este posibilă pentru funcții $G(u) = \varphi_c^{(k)}(u)$. Pentru orice k număr întreg, statistica $E(\varphi_c^{(k)}(\mathbf{w}^T\mathbf{v}))$ este numită moment gaussian al datelor.

Fie \mathbf{x} vectorul observațiilor perturbate și $\tilde{\mathbf{x}} = (\Sigma - \Sigma_\eta)^{-\frac{1}{2}}\mathbf{x}$. Estimarea componentelor independente ale modelului ICA cu zgomot

este realizată prin rezolvarea problemei variaționale,

$$(7) \max_{\|\mathbf{w}\|=1} [E(\varphi_d^{(k)}(\mathbf{w}^T \tilde{\mathbf{x}})) - E(\varphi_c^{(k)}(\mathbf{v}))]^2,$$

unde $d(\mathbf{w}) = \sqrt{c^2 - \mathbf{w}^T \tilde{\Sigma} \mathbf{w}}$. Calculul soluției problemei variaționale (7) poate fi realizat prin utilizarea unuia din algoritmi gradient corespunzători modelelor ICA fără zgomot.

În practică, algoritmi de estimare a componentelor independente ale modelului ICA cu zgomot sunt dezvoltați pentru $k = 0$, respectiv $k = -2$. În cazul $k = -2$, $\varphi^{(-2)}(u)$ este aproximat prin $G(u) = \frac{1}{2} \log \cosh(u)$.

Algoritmul FastICA pentru date perturbate

O variantă a algoritmului FastICA pentru estimarea componentelor independente ale modelelor cu zgomot este derivată pe baza funcției definite prin (6); regula de actualizare a algoritmului este,

$$(8) \mathbf{w}^* = E(\tilde{\mathbf{x}}g(\mathbf{w}^T \tilde{\mathbf{x}})) - (\mathbf{I} + \tilde{\Sigma})E(g'(\mathbf{w}^T \tilde{\mathbf{x}})),$$

$$\mathbf{w}^* = \frac{\mathbf{w}^*}{\|\mathbf{w}^*\|}, \text{ unde}$$

$\tilde{\Sigma} = E(\tilde{\mathbf{\eta}}\tilde{\mathbf{\eta}}^T) = (\Sigma - \Sigma_\mu)^{-\frac{1}{2}} \Sigma_\mu (\Sigma - \Sigma_\mu)^{-\frac{1}{2}}$ și g este derivata funcției G , deci poate fi selectată prin, $g(u) = \tanh(u)$, $g(u) = u \exp\left(-\frac{u^2}{2}\right)$, respectiv $g(u) = u^3$.

3. Tehnici de estimare a componentelor independente fără zgomot

În cadrul modelelor ICA cu zgomot definite de (1), nu este suficientă estimarea matricei mixtură. Într-adevăr, dacă $\mathbf{W} = \mathbf{A}^{-1}$, obținem

$$(10) \log L(\mathbf{A}, \mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(T)) = -\sum_{t=1}^T \left[\frac{1}{2} \|\mathbf{A}\mathbf{s}(t) - \mathbf{x}(t)\|_{\Sigma_\eta}^2 + \sum_{i=1}^n f_i(s_i(t)) \right] + C,$$

unde $\|\mathbf{m}\|_{\Sigma_\eta}^2 = \mathbf{m}^T \Sigma_\eta^{-1} \mathbf{m}$, $\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(T)$ sunt realizări ale componentelor independente, C este o constantă și, pentru $1 \leq i \leq n$, f_i este logaritmul densității de probabilitate a componentei independente s_i . Estimatorul

$\mathbf{W}\mathbf{x} = \mathbf{s} + \mathbf{W}\boldsymbol{\eta}$, deci estimări ale componentelor independente care înglobează componenta zgomot $\mathbf{W}\boldsymbol{\eta}$. Problema este de a estima componentele independente originale \mathbf{s} sau variante ale acestora care să conțină o cantitate minimă de zgomot. O variantă de rezolvare a problemei este obținută prin utilizarea estimațiilor MAP (Maximum A Posteriori). Fie $\boldsymbol{\theta}$ vectorul parametrilor; estimatorul MAP al lui $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{MAP}$, este definit prin valoarea vectorului $\boldsymbol{\theta}$ care maximizează densitatea posteriori a lui $\boldsymbol{\theta}$ pe baza măsurătorilor \mathbf{x}_T , $p_{\boldsymbol{\theta}/\mathbf{x}}(\boldsymbol{\theta}/\mathbf{x}_T)$. Estimatorul MAP este cea mai probabilă valoare a vectorului parametrilor $\boldsymbol{\theta}$ pentru datele observate \mathbf{x}_T . Densitatea posteriori $p_{\boldsymbol{\theta}/\mathbf{x}}(\boldsymbol{\theta}/\mathbf{x}_T)$ este calculată prin formula Bayes,

$$(9) p_{\boldsymbol{\theta}/\mathbf{x}}(\boldsymbol{\theta}/\mathbf{x}_T) = \frac{p_{\mathbf{x}/\boldsymbol{\theta}}(\mathbf{x}_T/\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{p_{\mathbf{x}}(\mathbf{x}_T)}.$$

Numitorul membrului drept al relației (9) nu depinde de $\boldsymbol{\theta}$, deci este suficientă determinarea vectorului $\boldsymbol{\theta}$ care maximizează $p_{\mathbf{x}/\boldsymbol{\theta}}(\mathbf{x}_T/\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, adică distribuția comună $p_{\boldsymbol{\theta},\mathbf{x}}(\boldsymbol{\theta}, \mathbf{x}_T)$. Estimatorul MAP $\hat{\boldsymbol{\theta}}_{MAP}$ este determinat rezolvând ecuația de verosimilitate maximă (în varianta logaritmată),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\boldsymbol{\theta}, \mathbf{x}_T) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\mathbf{x}_T/\boldsymbol{\theta}) + \frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\boldsymbol{\theta}) = 0$$

În scopul estimării componentele independente originale, este considerată funcția de verosimilitate,

MAP este obținut prin anularea gradientul funcției de verosimilitate (10) în raport cu $\mathbf{s}(t)$. Rezultă,

$$(11) \hat{\mathbf{A}}^T \Sigma_\eta^{-1} \hat{\mathbf{A}} \hat{\mathbf{s}}(t) - \hat{\mathbf{A}}^T \Sigma_\eta^{-1} \mathbf{x}(t) + f'(\hat{\mathbf{s}}(t)) = 0,$$

unde derivata log-densității de repartiție, f , este aplicată separat fiecărei componente a vectorului $\hat{\mathbf{s}}(t)$.

Calculul estimațiilor $\hat{\mathbf{s}}$ pe baza relației (11) este realizat de obicei prin optimizări numerice. Un caz particular este acela al zgomotului cu matricea de covarianță definită prin $\Sigma_\eta = \mathbf{A}\mathbf{A}^T\sigma^2$, corespunzător zgomotului de tip sursă. Din (11) obținem,

$$\sigma^2\hat{\mathbf{s}}(t) - \hat{\mathbf{A}}^{-1}\mathbf{x}(t) + f'(\hat{\mathbf{s}}(t)) = 0,$$

deci

$$(12) \hat{\mathbf{s}} = g(\hat{\mathbf{A}}^{-1}\mathbf{x}),$$

unde funcția scalară g este, [9]

$$(13) g^{-1}(u) = u + \sigma^2 f'(u).$$

Inversa funcției definite prin (13) nu poate fi de obicei obținută analitic. În continuare sunt prezentate câteva exemple în care g poate fi calculată [10].

Dacă s este distribuită Laplace (distribuția Laplace este supergaussiană), cu varianță 1, atunci,

$$p(s) = \frac{\exp(-\sqrt{2}|s|)}{\sqrt{2}}, \quad f'(s) = \sqrt{2} \operatorname{sign}(s) \text{ și}$$

$$(14) g(u) \cong \operatorname{sign}(u) \max(0, |u| - \sqrt{2}\sigma^2).$$

Funcția definită de (14) este o funcție de contracție.

$$(18) g(u) = \operatorname{sign}(u) \max\left(0, \frac{|u| - ad}{2} + \frac{1}{2} \sqrt{(|u| + ad)^2 - 4\sigma^2(\alpha + 3)}\right),$$

dacă $(|u| + ad)^2 - 4\sigma^2(\alpha + 3) \geq 0$, unde

$$a = \sqrt{\frac{\alpha(\alpha + 1)}{2}} \text{ și } g(u) = 0, \text{ în caz contrar.}$$

4. Eliminarea zgomotului prin contracția codului

Tehnica de estimarea componentelor independente originale ale modelului ICA cu zgomot poate fi extinsă pentru derivarea unui algoritm de eliminare a zgomotului dintr-un semnal dat. Fie observațiile perturbate, $\mathbf{x} = \mathbf{v} + \boldsymbol{\eta}$, unde \mathbf{v} este modelat prin,

$$(19) \mathbf{v} = \mathbf{A}\mathbf{s},$$

unde \mathbf{A} este matrice ortogonală și matricea de covarianță a vectorului zgomot este $\Sigma_\eta = \sigma^2\mathbf{I}$.

Dacă funcția scor f este aproximată ca o combinație a funcțiilor scor Laplace și gaussiană, $f'(s) = as + b \operatorname{sign}(s)$ cu $a, b > 0$, deci distribuția de probabilitate este,

$$(15) p(s) = C \exp\left(-\frac{as^2}{2} - b|s|\right), \quad C \text{ constantă,}$$

atunci,

$$(16) g(u) = \frac{1}{1 + \sigma^2 a} \operatorname{sign}(u) \max(0, |u| - b\sigma^2)$$

Funcția definită prin (15) este o funcție de contracție cu scalare adițională.

O variantă alternativă este utilizarea densității de probabilitate supergaussiene definite prin,

$$(17) p(s) = \frac{1}{2d} \frac{(\alpha + 2) \left(\frac{\alpha(\alpha + 1)}{2}\right)^{\frac{\alpha}{2} + 1}}{\sqrt{\frac{\alpha(\alpha + 1)}{2} + |s|}^{\alpha + 3}},$$

unde $\alpha, d > 0$. Pentru $\alpha \rightarrow \infty$, este obținută distribuția Laplace. Densitatea de probabilitate definită prin (17) are proprietatea că măsura kurtosis corespunzătoare este întotdeauna mai mare decât a distribuției Laplace și, pentru $\alpha \leq 2$, măsura kurtosis tinde la infinit. Funcția de contracție rezultată este,

În practică, matricea mixtură este aproximată prin intermediul unei matrice ortogonale (algoritm de ortogonalizare simetrică).

În scopul eliminării zgomotului datelor observate \mathbf{x} , putem proceda astfel. Sunt calculate estimațiile componentelor independente în vectorul $\hat{\mathbf{s}}$, utilizând estimatorul MAP. Datele sunt reconstruite prin, [10]

$$\hat{\mathbf{v}} = \hat{\mathbf{A}}\hat{\mathbf{s}}.$$

Metoda de eliminare a zgomotului presupune următoarele etape.

1. Modelul ICA definit prin (19) este estimat pe baza unui set de antrenare neperturbat \mathbf{v} . În continuare vom nota cu \mathbf{W}^T matricea mixtură ortogonală astfel obținută. Este estimat modelul densităților de probabilitate $p_i(s_i)$, $1 \leq i \leq n$, utilizând unul din modelele descrise prin(15), respectiv (17).

2. Pentru fiecare observație perturbată $\mathbf{x}(t)$, calculează $\mathbf{u}(t) = \mathbf{W}\mathbf{x}(t)$. Calculează $\hat{s}_i(t) = g_i(u_i(t))$, unde funcțiile g_i sunt definite prin (16), respectiv (18).
3. Calculează datele neperturbate $\hat{\mathbf{v}}(t) = \mathbf{W}^T \hat{\mathbf{s}}(t)$.

5. Aplicații ale modelelor ICA în procesarea imaginilor

Una din abordările fundamentale în procesarea de semnal este determinarea unui model statistic generativ pentru semnalul observat. Componentele sistemului generativ obținut constituie o modalitate de reprezentare a datelor, pe baza căreia pot fi dezvoltate tehnici de compresie, eliminare de zgomot și recunoaștere a formelor. De asemenea, acest tip de abordare permite modelarea proprietăților neuronilor în aria senzorilor primari.

Reprezentarea imaginilor este în general realizată pe baza unor transformări liniare în domeniul discret, aplicate datelor observate. Fie I o imagine monocromă; I poate fi reprezentată prin superpoziția liniară a unui set de caracteristici $a_i(x, y), 1 \leq i \leq n$,

$$(20) I(x, y) = \sum_{i=1}^n a_i(x, y) s_i,$$

unde, pentru $1 \leq i \leq n$, s_i este un coeficient stochastic. Mulțimea coeficienților stochastici este specifică fiecărui semnal imagistic. Alternativ, colectând pixelii imagine în vectorul $\mathbf{x} = (x_1, x_2, \dots, x_m)^T$, obținem reprezentarea ICA standard,

$$(21) \mathbf{x} = \mathbf{A}\mathbf{s}.$$

În cadrul reprezentării (21) a imaginii I colectată în vectorul \mathbf{x} , numărul componentelor independente este egal cu numărul variabilelor observate.

În aplicații, datorită dimensionalității semnalului imagistic observat, modelul de reprezentare (21) este utilizat pentru descrierea fiecărui bloc obținut prin partiționarea imaginii în ferestre (dimensiunea unei ferestre este aleasă de obicei 8×8).

Estimarea reprezentării definite prin relația (20) revine la calculul valorilor s_i și $a_i(x, y)$, pentru orice $1 \leq i \leq n$ și orice poziție

(x, y) a imaginii I . Determinarea parametrilor modelului (20) este realizată pe baza unui număr suficient de mare de observații asupra imaginii sau, în practică, ale blocurilor imaginii I .

În continuare vom considera că matricea \mathbf{A} este pătratică și inversabilă. Pentru orice $1 \leq i \leq n$ obținem, $s_i = \sum_{x,y} w_i(x, y) I(x, y)$,

unde w_i este filtrul invers și

$$(22) \mathbf{a}_i = \mathbf{A}\mathbf{A}^T \mathbf{w}_i = \mathbf{\Sigma} \mathbf{w}_i,$$

unde $\mathbf{\Sigma}$ este matricea de autocorelație a semnalului imagistic.

Relația (22) exprimă legătura dintre filtrele \mathbf{w}_i și vectorii bazei de reprezentare, \mathbf{a}_i . Vectorii bazei sunt obținuți prin filtrarea cu $\mathbf{\Sigma}$ a coeficienților vectorului \mathbf{w}_i . În general, în cazul datelor imagistice, matricea de autocorelație este simetrică și reprezintă un filtru trece-jos, deci vectorii \mathbf{a}_i sunt variante nivelate ale filtrelor \mathbf{w}_i .

Problema este de a determina principiile estimării transformării pe baza datelor observate. Punctul de plecare în soluționarea problemei este un principiu de reprezentare numit *codificare rară*. În cadrul acestui tip de reprezentare, datele observate sunt modelate prin intermediul unei baze de vectori astfel încât, la un moment de timp, numai un număr mic de vectori bază sunt "activi". În interpretare neuronală, fiecare vector al bazei corespunde unui neuron și coeficienții s_i sunt obținuți prin activarea sistemului de neuroni. Rezultă că, pentru un bloc dat, sunt activați un număr mic de neuroni. O modalitate echivalentă de exprimare a principiului codificării rare este: un neuron dat este activat rar. Cu alte cuvinte, coeficienții s_i au o distribuție rară (de tip supergaussian).

În ipoteza în care codul rar este liniar, transformarea aplicată datelor observate este estimată prin formularea unei măsurii a proprietății componentelor s_i de a avea o distribuție rară și maximizarea acelei măsurii. Deoarece distribuțiile rare sunt legate de proprietatea de supergaussianitate, pot fi folosite măsurile kurtosis, respectiv negentropia, deci sunt

utilizați algoritmi de estimare a componentelor independente ale modelelor ICA fără zgomot. Rezultă că principiul codificării rare poate fi considerat ca o metodă ICA. [9]

Utilitatea principiului de codificare rară este remarcată în aplicații de compresie și eliminare de zgomot în cazul semnalelor imagistice. În problemele de compresie, deoarece, pentru un set de date observate, relativ puține componente sunt nenule, codificarea datelor poate fi realizată eficient prin considerarea componentelor nenule. Pentru eliminarea zgomotului, pot fi utilizate proceduri de tip prag care determină componentele active și setează pe 0 celelalte componente (observațiile reprezintă zgomot).

6. Eliminarea zgomotului repartizat $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ utilizând PCA și contracția codului

Modelul propus pentru eliminarea zgomotului $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ utilizează tehnici de tip PCA și contracție a codului [21]. În continuare vom considera că setul imaginilor originale \mathbf{X}_0 este perturbat aditiv cu zgomotul $\boldsymbol{\eta} = (\boldsymbol{\eta}_t, t \geq 0)$, modelat prin intermediul unui proces stocastic staționar: pentru orice t , $\boldsymbol{\eta}_t$ este distribuit $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, cu σ^2 cunoscut. De asemenea, presupunem că \mathbf{X}_0 este proces staționar, cu $\boldsymbol{\mu} = E(\mathbf{X}_0(t))$ și $\boldsymbol{\Sigma} = Cov(\mathbf{X}_0(t), (\mathbf{X}_0(t))^T)$ cunoscute. Setul de imagini observate este, $\mathbf{X} = \mathbf{X}_0 + \boldsymbol{\eta}$.

Datele observate reprezintă imagini monocrome și sunt prelucrate considerând descompunerea în blocuri de dimensiune 8×8 , ulterior liniarizate ($n = 64$). În etapa de preprocesare datele sunt reduse la intervalul $[0,1]$ (rezultă $0 < \sigma^2 < 1$) și centrate,

$$\mathbf{Y} = \mathbf{X} - E(\mathbf{X}) = \mathbf{X}_0 - \boldsymbol{\mu} + \boldsymbol{\eta}$$

Rezultă, $Cov(\mathbf{Y}, \mathbf{Y}^T) = \boldsymbol{\Sigma} + \sigma^2 \mathbf{I}_n$.

Fie $\mathbf{A} = \boldsymbol{\Phi} \boldsymbol{\Lambda}^{-\frac{1}{2}}$, unde $\boldsymbol{\Phi}$ este matricea ortonormală cu coloane vectorii proprii ai lui $\boldsymbol{\Sigma}$ și $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, $\lambda_i = 1 + \frac{\sigma^2}{\theta_i}$,

unde $\theta_1, \theta_2, \dots, \theta_n$ sunt valorile proprii ale matricei $\boldsymbol{\Sigma}$.

Rezultă că fiecare coloană din \mathbf{A} corespunde unui vector propriu al matricei $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma} + \sigma^2 \mathbf{I}_n)$, iar $\lambda_1, \lambda_2, \dots, \lambda_n$ sunt valorile proprii asociate. Matricea \mathbf{A} diagonalizează simultan matricele $\boldsymbol{\Sigma}$ și $\boldsymbol{\Sigma} + \sigma^2 \mathbf{I}_n$, $\mathbf{A}^T \boldsymbol{\Sigma} \mathbf{A} = \mathbf{I}_n$, $\mathbf{A}^T (\boldsymbol{\Sigma} + \sigma^2 \mathbf{I}_n) \mathbf{A} = \boldsymbol{\Lambda}$

Fie transformarea,

$$(23) \mathbf{Z} = \mathbf{A}^T \mathbf{Y} = \mathbf{A}^T (\mathbf{X}_0 - \boldsymbol{\mu}) + \mathbf{A}^T \boldsymbol{\eta}.$$

Obținem, $Cov(\mathbf{A}^T \boldsymbol{\eta}, (\mathbf{A}^T \boldsymbol{\eta})^T) = \sigma^2 \boldsymbol{\Lambda}^{-1}$

Deoarece $\boldsymbol{\eta}_t$ este distribuit $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, vectorul $\mathbf{A}^T \boldsymbol{\eta}$ are distribuția $N(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}^{-1})$, deci elementele componentei zgomot rezultate, $\mathbf{A}^T \boldsymbol{\eta}$, sunt independente. În continuare vectorului \mathbf{Z} îi este aplicată funcția de contracție

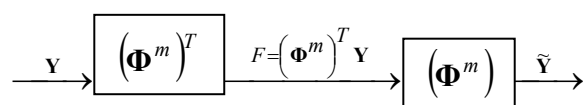
$$g(u) = \text{sign}(u) \max\left(0, |u| - \sqrt{2} \frac{\sigma^2}{\lambda_i}\right),$$

și rezultă o aproximare a lui \mathbf{Z} în varianta fără zgomot,

$$(24) \mathbf{Z}_0 = \mathbf{A}^T (\mathbf{X}_0 - \boldsymbol{\mu}).$$

Deoarece $\mathbf{A} \mathbf{A}^T = \boldsymbol{\Sigma}^{-1}$ și din relațiile (23) și (24) este obținută o aproximare a setului de imagini inițiale, $\hat{\mathbf{X}}_0 = \boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{A} \mathbf{Z}_0$.

Tehnica propusă mai sus, CSPCA, este combinată cu o schemă de compresie/decompresie, astfel încât procesul de eliminare a zgomotului este tratat în spațiul caracteristicilor principale. Fie $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \dots, \boldsymbol{\Phi}_n$ vectorii proprii unitari ai matricei $\boldsymbol{\Sigma}$ și $\theta_1 \geq \theta_2 \geq \dots \geq \theta_n$ valorile proprii corespunzătoare. Pentru orice $1 \leq m \leq n$, fie $\boldsymbol{\Phi}^m = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \dots, \boldsymbol{\Phi}_m)$ și $\boldsymbol{\Lambda}_m = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$. Schema de compresie/decompresie optimală LMS este,



unde $\tilde{\mathbf{Y}} = \boldsymbol{\Phi}^m (\boldsymbol{\Phi}^m)^T \mathbf{Y}$.

Modulul de eliminare a zgomotului prezentat mai sus este implementat în spațiul m -dimensional al caracteristicilor principale, F . Obținem următoarea schemă de compresie/decompresie și restaurare:

1. Compresia

$$\xrightarrow{Y=X-\mu} (\Lambda_m)^{-\frac{1}{2}} (\Phi^m)^T \rightarrow F = (\Lambda_m)^{-\frac{1}{2}} (\Phi^m)^T Y$$

2. Modulul de eliminare a zgomotului



3. Decompresia

$$\xrightarrow{F_0} \left((\Lambda_m)^{-\frac{1}{2}} (\Phi^m)^T \right)^+ \rightarrow \hat{X},$$

unde \hat{X} este imaginea restaurată,

$$(25) \hat{X} = \left((\Lambda_m)^{-\frac{1}{2}} (\Phi^m)^T \right)^+ F_0.$$

Bibliografie

[1] S. Bannour, M.R. Azimi-Sadjadi, "Principal Component Extraction Using Recursive Least Squares Learning" IEEE Transaction on Neural Networks, vol.6,no.2, 1995

[2] C. Chatterjee, V.P. Roychowdhury, E.K.P. Chong, "On Relative Convergence Properties of Principal Component Analysis Algorithms", IEEE Transaction on Neural Networks, vol.9,no.2, 1998

[3] C. Cocianu, L. State, V. Ștefănescu, P. Vlamos, "On the Efficiency of a Certain Class of Noise Removal Algorithms in Solving Image Processing Tasks", Proceedings of ICINCO 2004, Setubal, Portugal

[4] C. Cocianu, L. State, P. Vlamos, V. Ștefănescu, "PCA Based Shrinkage Attempt to Noise Removal", Proceedings of 35th CIE, August 2005, Istanbul, Turcia, în curs de apariție

[5] K.I. Diamantaras, S.Y. Kung, Principal Component Neural Networks: theory and applications, John Wiley & Sons, 1996

[6] J. Karhunen, E. Oja, "New Methods for Stochastic Approximations of Truncated Karhunen-Loeve Expansions", Proceedings 6th International Conference on Pattern Recognition, Springer Verlag, 1982

[7] H.J. Kushner., D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, Springer Verlag, 1978

[8] S. Haykin, Neural Networks A Comprehensive

Foundation, Prentice Hall, Inc. 1999

[9] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, 2001

[10] A. Hyvarinen, P. Hoyer, E. Oja, "Image Denoising by Sparse Code Shrinkage", www.cis.hut.fi/projects/ica, November, 1999

[11] J. Mao, A.K. Jain, "Artificial Neural Networks for Feature Extraction and Multivariate Data Projection", IEEE Transaction on Neural Networks, vol.6,no.2, 1995

[12] K. Matsuoka, M. Kawamoto, "A Neural Network that Self-Organizes to Perform Three Operations Related to Principal Component Analysis", Neural Networks, vol.7,no.5, 1994

[13] E. Oja, "Principal Components, Minor Components and Linear Neural Networks", Neural Networks, vol. 5, 1992

[14] M.D. Plumbley, "Lyapunov Functions for Convergence of Principal Component Algorithms", Neural Networks, vol.8,no.1, 1995

[15] T.D. Sanger, "An Optimality Principle for Unsupervised Learning", Advances in Neural Information Systems, ed. D.S. Touretzky, Morgan Kaufmann, 1989

[16] L. State, C. Cocianu, P. Vlamos, "Attempts in Using Statistical Tools for Image Restoration Purposes", Proceedings of SCI2001, Orlando, USA, July 22-25, 2001.

[17] L. State, C. Cocianu, P. Vlamos, "A Regressive Technique of Image Restorations", Proceedings of the 29th ICC&IE, Nov. 1-3, 2001, Montreal, Canada

[18] L. State, C. Cocianu, P. Vlamos, "A Multiresolution Based Approach of Data Compression/ Decompression", Proceedings of the 29th ICC&IE, Nov. 1-3, 2001, Montreal, Canada

[19] L. State, C. Cocianu, V. Ștefănescu, P. Vlamos, "Neural Network for Principal Component Analysis with Applications in Image Compression", Proceedings of CCCT 2003, Texas, USA, 2003

[20] L. State, C. Cocianu, V. Ștefănescu, P. Vlamos, "A Comparative Analysis on the Performance of a certain class of PCA Algorithms in solving Image Processing tasks", Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2003), Orlando, USA, July 27-30, 2003

[21] L. State, C. Cocianu, P. Vlamos, V. Ștefănescu, "Noise Removal Algorithm Based on Code Shrinkage Techniques", Proceedings of the 9th World Multiconference on Systemics, Cybernetics and Informatics (WMSCII 2005), Orlando, USA, July 2005.