

Αλγοριθμικές Μέθοδοι Βελτιστοποίησης με Έμφαση σε Κατανεμημένα Προβλήματα

Εισαγωγή στην Εκμάθηση Λεξικών

Δημήτρης Αμπελιώτης

Επίκουρος Καθηγητής, Ιόνιο Πανεπιστήμιο

Περιεχόμενα

- Αραιές Αναπαραστάσεις (Διαφάνειες Δρ. Χ. Μαυροκεφαλίδη)
 - Εισαγωγή
 - Υποκαθορισμένα συστήματα γραμμικών εξισώσεων
 - Εύρεση αραιών λύσεων
 - Ο αλγόριθμος OMP
- Εκμάθηση Λεξικού (ΕΛ) και Αλγόριθμοι
 - Εισαγωγή
 - Αλγόριθμοι Αραιής Αναπαράστασης
 - Αλγόριθμοι Ανανέωσης Λεξικών
- Εφαρμογές της Εκμάθησης Λεξικού
 - Συμπύεση / Απομάκρυνση Θορύβου
 - Συζευγμένη ΕΛ – Super resolution
 - Discriminative DL – Classification
 - Non-uniform noise - Inpainting
- Διάφορες άλλες μορφές ΕΛ
 - Συνελικτική Εκμάθηση Λεξικού
 - Online/Distributed μορφές

Εισαγωγή (1/7)

- Οι αραιές αναπαραστάσεις σημάτων αποτελούν μια σημαντική προσέγγιση για την **μοντελοποίηση** σημάτων και μελετώνται για διάφορους λόγους.
 - Επειδή βελτιώνεται η απόδοση των τεχνικών στατιστικής επεξεργασίας σήματος και μάθησης.
 - Επειδή, στην φύση, τα παραγόμενα φυσικά σήματα τείνουν να επιδέχονται μοντελοποίησης με φειδωλές (περιεκτικές) αναπαραστάσεις.
- Το παράδειγμα της μοντελοποίησης της κρουστικής απόκρισης ακουστικών και τηλεπικοινωνιακών μονοπατιών.



Εισαγωγή (2/7)

- Το παράδειγμα της συμπίεσης εικόνας:
 - Ο μετασχηματισμός της εικόνας σε έναν νέο, κατάλληλο χώρο επιτρέπει την αραιή αναπαράστασή της.
 - Αν Φ ($l \times l$) είναι ο πίνακας που περιγράφει έναν τέτοιο γραμμικό μετασχηματισμό (π.χ., διακριτού συνημιτόνου, DCT), τότε η νέα αναπαράσταση i_s της εικόνας i προκύπτει ως:

$$i_s = \Phi^H i \Rightarrow i = \Phi i_s, \quad \Phi^H \Phi = I$$

- Συνήθως, στο i_s , τα σημαντικά στοιχεία είναι λίγα σε αριθμό σε σύγκριση με το σύνολο των στοιχείων.
- Για την συμπίεση, διατηρούνται μόνο $k \ll l$ σημαντικά στοιχεία και τα υπόλοιπα τίθενται στο μηδέν, δημιουργώντας το αραιό διάνυσμα \hat{i}_s .

Εισαγωγή (3/7)

- ...συνέχεια.
 - Για την αποσυμπίεση, χρησιμοποιείται ο αντίστροφος μετασχηματισμός:

$$\hat{\mathbf{i}} = \mathbf{\Phi} \hat{\mathbf{i}}_s$$

- Το παραπάνω μπορεί να γραφεί εναλλακτικά ως

$$\hat{\mathbf{i}} = \mathbf{\Phi} \hat{\mathbf{i}}_s = \sum_{j=1}^k \hat{\mathbf{i}}_s(t_j) \boldsymbol{\varphi}_{t_j}$$

- Όπου, $t_j, j = 1, 2, \dots, k$, είναι οι θέσεις των σημαντικών στοιχείων, $\hat{\mathbf{i}}_s(t_j)$ η τιμή τους και $\boldsymbol{\varphi}_{t_j}$ η αντίστοιχη στήλη του πίνακα $\mathbf{\Phi}$.
- Το εν λόγω άθροισμα αποτελεί μια αραιή αναπαράσταση του σήματος $\hat{\mathbf{i}}$.

Εισαγωγή (4/7)

- ...συνέχεια
 - Στο προηγούμενο, θεωρήθηκε πως η αρχική εικόνα έχει ληφθεί κανονικά. Δηλαδή, αν και καταλήξαμε με k σημαντικά στοιχεία, χρειάστηκε να συλλέξουμε καταρχήν $l \gg k$ μετρήσεις.
 - Υπάρχει εναλλακτική προσέγγιση όπου να βασιζόμαστε εξαρχής μόνο σε $N = O(k) \ll l$ μετρήσεις για να καταλήξουμε στα k στοιχεία;
 - Η απάντηση στο παραπάνω ερώτημα έχει οδηγήσει στην έννοια του compressive sampling (εδώ και τουλάχιστον μια δεκαετία), η οποία καταλήγει στην απαίτηση επίλυσης ενός υποκαθορισμένου συστήματος γραμμικών εξισώσεων.

Εισαγωγή (5/7)

- ...συνέχεια
 - Έστω ένα σύστημα λήψης εικόνων που περιγράφεται από τον πίνακα μέτρησης A ($N \times l$). Τότε, θα μπορούσαμε να συλλέξουμε τις μετρήσεις y ως:
$$y = Ai = A\Phi i_s$$
 - Στην περίπτωση αυτή, χρησιμοποιούμε τις μετρήσεις y για να υπολογίσουμε το αραιό διάνυσμα i_s .
 - Οπότε, η εικόνα i μπορεί να ανακατασκευαστεί αφού γνωρίζουμε τον πίνακα Φ .

Εισαγωγή (6/7)

- Προηγουμένως, είδαμε ένα παράδειγμα αραιής αναπαράστασης που βασίζεται σε έναν ορθοκανονικό πίνακα Φ .
 - Αυτοί οι πίνακες, αν και δουλεύουν αποδοτικά για ορισμένα σήματα, είναι περιοριστικοί σε ορισμένες περιπτώσεις.
- Αν θέλουμε να καθορίσουμε μια αραιή αναπαράσταση για ένα σήμα μουσικής m ($l \times 1$);
 - Το σήμα λόγω των οργάνων έχει αρμονικές.
 - Το σήμα φωνής του καλλιτέχνη είναι πιο περίπλοκο.
- Μια εικόνα αποτελείται από ομαλές περιοχές και περιοχές με έντονες ακμές.

Εισαγωγή (7/7)

- Για πιο περίπλοκα σήματα, οι αναπαραστάσεις δύναται να βασίζονται σε μη ορθογώνιους (overcomplete) μετασχηματισμούς.
- Για παράδειγμα, ένα σήμα m θα μπορούσε να αναπαρασταθεί από έναν νέο πίνακα Φ ($l \times 2l$) που αποτελείται από έναν πίνακα DFT και έναν πίνακα DWT.
$$m = \Phi m_s$$
 - Όπου το παραπάνω σύστημα είναι υποκαθορισμένο και «έρχεται» πιο κοντά στο μοντέλο του compressive sampling που είδαμε προηγουμένως.
- Ο πίνακας Φ , στην γενική περίπτωση, ονομάζεται λεξικό, έχει περισσότερες στήλες και δύναται να χρησιμοποιηθεί για την σύνθεση μιας μεγάλης ποικιλίας σημάτων καθορίζοντας μια αραιή αναπαράσταση.

Σύστημα γραμμικών εξισώσεων (1/2)

- Το πρόβλημα της επίλυσης ενός συστήματος γραμμικών εξισώσεων είναι κεντρικής σημασίας σε διάφορες επιστημονικές περιοχές.
- Η βασική μορφή ενός συστήματος γραμμικών εξισώσεων είναι:
$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}$$
 - $\mathbf{y} \in \mathbb{R}^N$: σύνολο μετρήσεων ή παρατηρήσεων.
 - $\boldsymbol{\theta} \in \mathbb{R}^l$: άγνωστοι παράμετροι.
 - $\mathbf{X} \in \mathbb{R}^{N \times l}$: γνωστός πίνακας που καθορίζει το γραμμικό μοντέλο του «φαινομένου» που θέλουμε να μελετήσουμε.
- Μέσω της γραμμικής άλγεβρας, η επίλυση έχει μελετηθεί διεξοδικά από διάφορες πλευρές.
 - Ύπαρξη μοναδικής λύσης
 - Ευρωστία λύσης
 - Υπολογιστική πολυπλοκότητα

Σύστημα γραμμικών εξισώσεων (2/2)

- Το σύστημα $y = X\theta$, ανάλογα με τις παραμέτρους l, N μπορεί να έχει τις ακόλουθες μορφές:

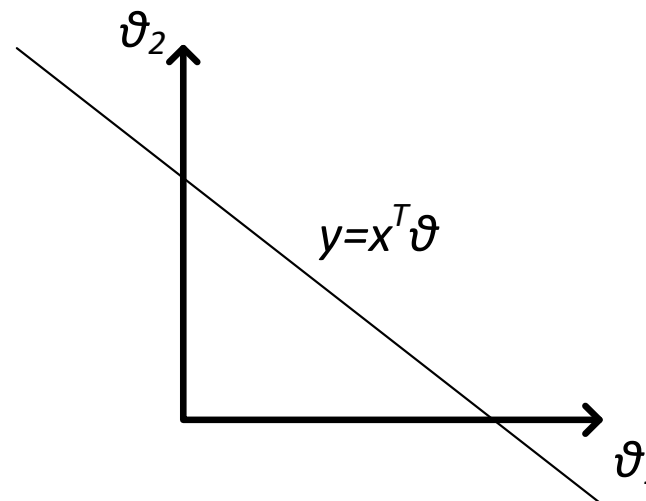
$$\begin{array}{c} y \\ | \\ | \\ | \end{array} = \begin{array}{c} \boxed{X} \\ N=l \end{array} \begin{array}{c} \theta \\ | \\ | \\ | \end{array}$$

$$\begin{array}{c} y \\ | \\ | \\ | \\ | \end{array} = \begin{array}{c} \boxed{X} \\ N>l \end{array} \begin{array}{c} \theta \\ | \\ | \end{array}$$

$$\begin{array}{c} y \\ | \\ | \\ | \end{array} = \begin{array}{c} \boxed{X} \\ N<l \end{array} \begin{array}{c} \theta \\ | \\ | \\ | \\ | \end{array}$$

Υποκαθορισμένο σ.γ. εξισώσεων (1/3)

- Στην συνέχεια θα επικεντρωθούμε στην περίπτωση $N < l$, και θα θεωρήσουμε πως $rank(\mathbf{X}) = N$.
 - Το σύστημα γραμμικών εξισώσεων ονομάζεται σε αυτή την περίπτωση υποκαθορισμένο (underdetermined).
- Ένα υποκαθορισμένο σύστημα έχει άπειρες λύσεις, οι οποίες βρίσκονται στην τομή των υπερεπιπέδων που καθορίζονται από τις εξισώσεις του συστήματος (η τομή γιατί υπάρχει;).
- Για παράδειγμα, εάν $l = 2$ (και, άρα, $N = 1$), τότε οι λύσεις «βρίσκονται» πάνω σε μια ευθεία (ποια η σχέση με τον μηδενικό χώρο;).



Υποκαθορισμένο σ.γ. εξισώσεων (2/3)

- Για παράδειγμα, εάν $l = 3$ (και $N = 2$), τότε οι λύσεις «βρίσκονται» στην τομή δυο επιπέδων.

ϑ_2

$$y_2 = x_2^T \vartheta$$

$$y_1 = x_1^T \vartheta$$

$$y = X\vartheta$$

ϑ_1

ϑ_3

- Το ερώτημα που καλούμαστε να απαντήσουμε είναι «Ποια λύση επιλέγουμε από αυτό το άπειρο σύνολο λύσεων».

Υποκαθορισμένο σ.γ. εξισώσεων (3/3)

- Για την επίλυση ενός υποκαθορισμένου συστήματος γραμμικών εξισώσεων, απαιτείται ο καθορισμός επιπλέον κριτηρίων που θα πρέπει να ικανοποιούνται από την λύση.
- Μια βασική προσέγγιση στηρίζεται στην χρήση μιας συνάρτησης κόστους $J(\boldsymbol{\theta})$ για την «**διαβάθμιση**» των άπειρων λύσεων και την επιλογή της πιο κατάλληλης.
- Το γενικό πρόβλημα γράφεται ως εξής:

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}), \quad s. t. \mathbf{y} = \mathbf{X}\boldsymbol{\theta}$$

- Συνηθισμένες επιλογές για την συνάρτηση $J(\boldsymbol{\theta})$:
 - $\|\boldsymbol{\theta}\|_0, \|\boldsymbol{\theta}\|_1, \|\boldsymbol{\theta}\|_2^2$

Η περίπτωση της $\|\boldsymbol{\theta}\|_2^2$

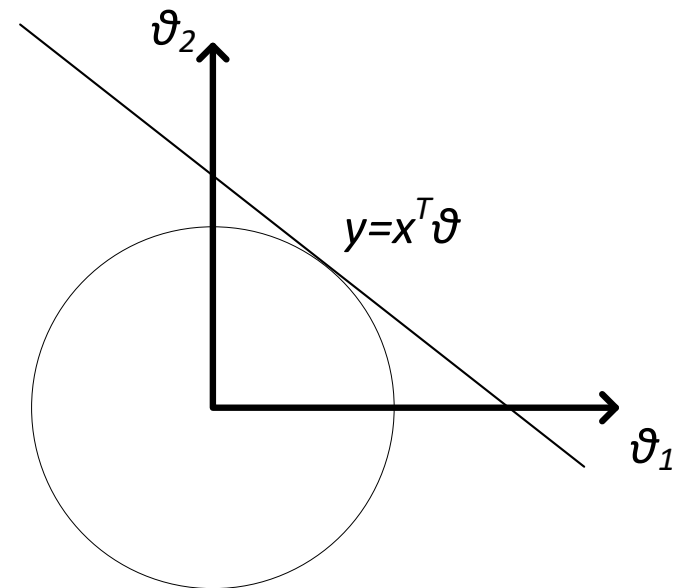
- Το πρόβλημα εξειδικεύεται ως

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_2^2, \quad s. t. \mathbf{y} = \mathbf{X}\boldsymbol{\theta}$$

- Η λύση με χρήση πολλαπλασιαστών Lagrange δίνεται ως:

$$\hat{\boldsymbol{\theta}} = \mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{y}$$

- Γεωμετρικά, η λύση περιγράφεται όπως φαίνεται στο σχήμα.
- Τι πρόβλημα υπάρχει στην περίπτωση που αναζητούμε ένα αραιό διάνυσμα $\boldsymbol{\theta}$;

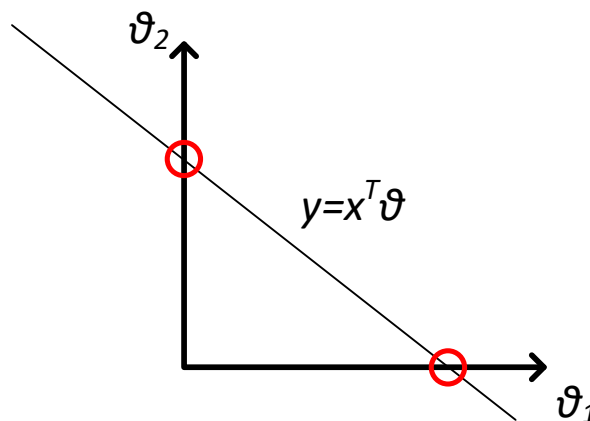


Η περίπτωση της $\|\theta\|_0$

- Το πρόβλημα εξειδικεύεται ως

$$\min_{\theta} \|\theta\|_0, \quad s.t. \mathbf{y} = \mathbf{X}\theta$$

- Το πρόβλημα βελτιστοποίησης έχει συνδυαστική πολυπλοκότητα για την εύρεση της βέλτιστης λύσης.
 - Στην πράξη, επιλύεται με προσεγγιστικούς αλγορίθμους (greedy algorithms όπως τον Orthogonal Matching Pursuit)
- Το πρόβλημα δεν έχει πάντα μοναδική λύση.



Η περίπτωση της $\|\boldsymbol{\theta}\|_1$ (1/2)

- Το πρόβλημα εξειδικεύεται ως

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1, \quad s. t. \mathbf{y} = \mathbf{X}\boldsymbol{\theta}$$

- Το παραπάνω πρόβλημα είναι γνωστό και ως Basis Pursuit.
- Η επίλυση το προβλήματος βελτιστοποίησης συνδέεται με την μετατροπή του σε ένα πρόβλημα γραμμικού προγραμματισμού.
 - Αυτά τα προβλήματα λύνονται αποδοτικά με διάφορες μεθόδους όπως την simplex ή την interior point.

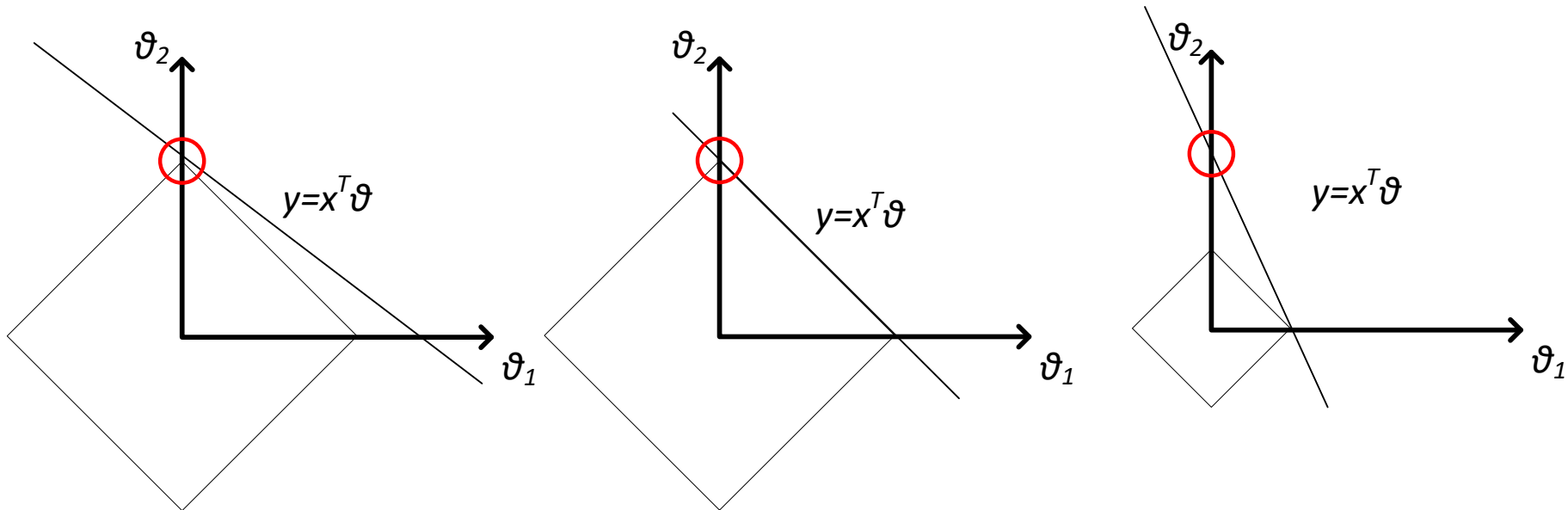
- Η μορφή ενός προβλήματος γραμμικού προγραμματισμού είναι:

$$\min_{\mathbf{x}} \mathbf{c}^T \mathbf{x}, \quad s. t. \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

- Πως γίνεται η μετατροπή του προβλήματος σε πρόβλημα γραμμικού προγραμματισμού;

Η περίπτωση της $\|\theta\|_1$ (2/2)

- Γεωμετρικά, η λύση περιγράφεται ως:



- Η επιλογή (ή ο σχεδιασμός) του πίνακα X είναι κρίσιμης σημασίας για την επίτευξη μοναδικής λύσης καθώς και της σωστής λύσης.

Ο αλγόριθμος OMP (1/4)

- Στην βιβλιογραφία έχουν εμφανιστεί διάφοροι αποδοτικοί αλγόριθμοι για την επίλυση των προβλημάτων που έχουν παρουσιαστεί.
 - Εφαρμόζονται είτε σε γενικούς πίνακες X είτε σε ειδικότερους, εκμεταλλευόμενοι την ιδιαίτερη δομή που μπορεί να έχουν.
 - Βασίζονται σε ευρετικές / προσεγγιστικές διαδικασίες, σε διαδικασίες κυρτού προγραμματισμού, κ.λπ.
- Στην συνέχεια θα επικεντρωθούμε σε έναν ευρετικό (greedy) αλγόριθμο.
 - Γενικότερα, οι αλγόριθμοι αυτού του τύπου προσπαθούν να εκτιμήσουν τις θέσεις και τις τιμές των μη μηδενικών συντελεστών του διανύσματος θ (λύνουν επί της ουσίας το πρόβλημα όταν χρησιμοποιείται $J(\theta) = \|\theta\|_0$).
 - Αυτό επιτυγχάνεται από μια σειρά απλών βημάτων.

Ο αλγόριθμος OMP (2/4)

- Ο αλγόριθμος Orthogonal Matching Pursuit (OMP) λαμβάνει ως **είσοδο** τον πίνακα X , το διάνυσμα y και μια παράμετρο k , που καθορίζει το πλήθος των μη μηδενικών στοιχείων που θα καθοριστούν στο διάνυσμα θ .
- Ο OMP επιστρέφει ως έξοδο το διάνυσμα $\hat{\theta}$ που αποτελείται από k μη μηδενικά στοιχεία.
- Αρχικά, ορίζονται τα εξής:
 - $S = \emptyset$: Σύνολο που περιέχει τις θέσεις των μη μηδενικών στοιχείων του διανύσματος θ .
 - $e = y$: Το σφάλμα αναπαράστασης.
 - $\theta = \mathbf{0}$

Ο αλγόριθμος OMP (3/4)

- Ο αλγόριθμος OMP επαναλαμβάνει τα ακόλουθα βήματα:
 1. $\hat{j} = \operatorname{argmax}_j |\mathbf{x}_j^T \mathbf{e}| / \|\mathbf{x}_j\|_2$ (βρες την στήλη του πίνακα \mathbf{X} που είναι πιο συσχετισμένη με το τρέχον σφάλμα).
 2. $S = S \cup \{\hat{j}\}$ (ανανέωσε το σύνολο με τις θέσεις μη μηδενικών στοιχείων του διανύσματος $\boldsymbol{\theta}$).
 3. $\boldsymbol{\theta}_S = \mathbf{X}_S^+ \mathbf{y}$ (υπολόγισε, υπό την έννοια των ελαχίστων τετραγώνων, τις τρέχουσες τιμές των μη μηδενικών στοιχείων).
 - \mathbf{X}_S : πίνακας ($N \times |S|$) που περιέχει τις έως τώρα επιλεγμένες στήλες του πίνακα \mathbf{X} .
 4. $\mathbf{e} = \mathbf{y} - \mathbf{X}_S \boldsymbol{\theta}_S$ (ανανέωσε το σφάλμα αναπαράστασης).
 5. Επανάλαβε τα παραπάνω για k φορές.

Ο αλγόριθμος OMP (4/4)

- Μερικές παρατηρήσεις για τον OMP είναι οι ακόλουθες:
 - Γενικώς, ο αλγόριθμος δεν εγγυάται πως θα καταλήξει κοντά στην πραγματική λύση.
 - Το σφάλμα e είναι κάθετο στις στήλες που έχουν επιλεχθεί σε κάθε επανάληψη (έτσι μια στήλη μπορεί να επιλεγεί μόνο μια φορά).
 - Η επιλογή της στήλης με την μέγιστη συσχέτιση με το τρέχον σφάλμα έχει ως αποτέλεσμα την μέγιστη ελαχιστοποίηση του δεύτερου.

Βιβλιογραφία

- S. Theodoridis, “Machine Learning: A Bayesian and Optimization Perspective”, Academic Press, 2015.

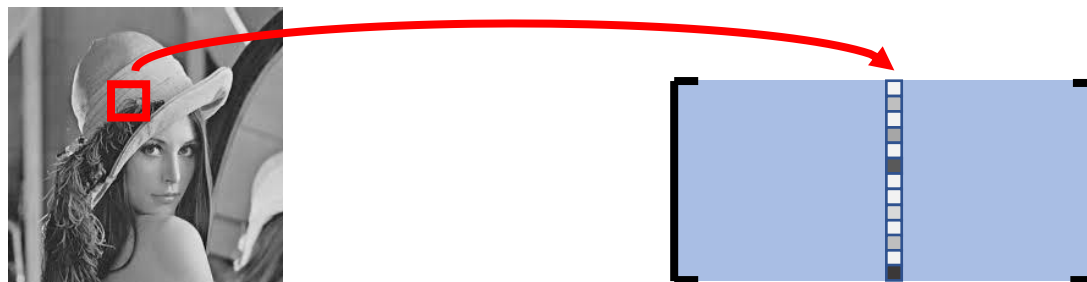
Εκμάθηση Λεξικού και Αλγόριθμοι

Εισαγωγή

- The emergence of **compressed sensing** extended the notion of bandlimited signals to signals that are sparse in **some suitable domain**, other than frequency
- Furthermore, it has been shown that **many physical signals** are indeed sparse
- DL constitutes a **data-driven** approach to **learn** a domain in which the signals of interest are sparse
- DL resembles orthogonal transforms such as the Fourier transform, but orthogonality is sacrificed for the sake of sparsity

Εισαγωγή

- Consider a **data matrix** $Y \in \mathbb{R}^{m \times n}$ (m : dimension, n : # of samples)
- Example (a): Image patches (overlapping or not)

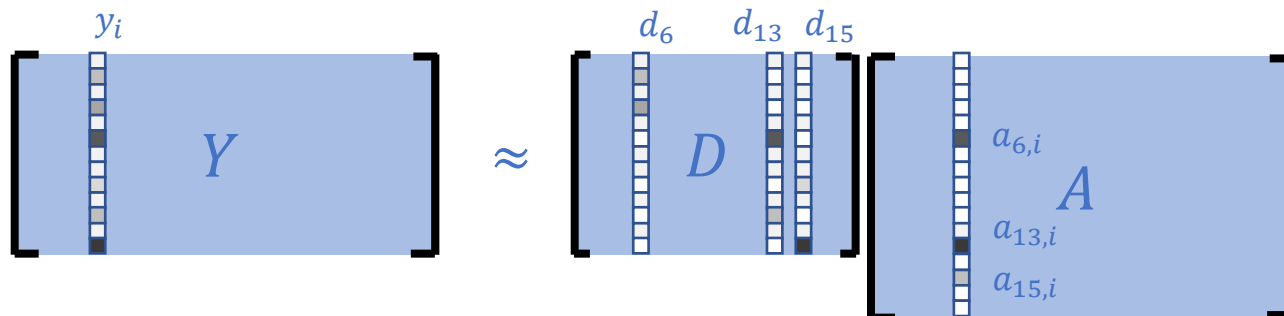


- Example (b): Spectral response



What is Dictionary Learning?

- The scope is to compute some matrix $D \in \mathbb{R}^{m \times k}$, so that $Y \approx D \cdot A$



- Furthermore, matrix $A \in \mathbb{R}^{k \times n}$ is required to satisfy some sparsity criterion (e.g., fixed number of non-zero elements at each column)
- Thus, each (vector) sample y_i can be expressed as a linear combination of a few columns of the **dictionary** matrix D , e.g.,

$$y_i \approx a_{6,i} \cdot d_6 + a_{13,i} \cdot d_{13} + a_{15,i} \cdot d_{15}$$

What is Dictionary Learning?

- In most cases, DL follows from the minimization (or, approximate minimization) of a cost function of the form:

$$C(D, A) = F(D, A) + \lambda \cdot S(A)$$

where $F(D, A)$ is some **fidelity cost** and $S(A)$ is some **sparsity cost** for matrix A . The parameter λ controls the relative importance of the two terms

- For example:
 - $F(D, A) = \|Y - DA\|_F^2$ (Frobenius norm)
 - $S(A) = \|A\|_0$ (number of non-zero elements, combinatorial optimization)
 - $S(A) = \|A\|_1$ (sum of absolute values of elements, Lasso)

What is Dictionary Learning?

- The cost function

$$C(D, A) = F(D, A) + \lambda \cdot S(A)$$

for various practical $F(D, A)$ and $S(A)$ is not convex, but it is **bi-convex**:

- Convex for D , when A is fixed
- Convex for A , when D is fixed
- Thus, the dominant approach for optimization is the so-called **alternating optimization** approach:
 - **Sparse Approximation step**: Optimize for A_{t+1} with fixed D_t
 - **Dictionary Update step**: Optimize for D_{t+1} with fixed A_{t+1}
- The dictionary atoms may require normalization after each iteration to have unit norm

Sparse Approximation Algorithms

- Choices of algorithms for ***Sparse Approximation***:
 - Matching Pursuit¹ (MP)
 - Orthogonal Matching Pursuit² (OMP)
 - Focal Underdetermined System Solver³ (FOCUSS)
 - Iterated Soft Thresholding⁴ (ISTA)
 - Fast Iterated Soft Thresholding⁵ (FISTA)
 - ...

¹ S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” IEEE Transactions on signal processing, vol. 41, no. 12, pp. 3397–3415, 1993.

² J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” IEEE Transactions on Information theory, vol. 50, no. 10, pp. 2231–2242, 2004.

³ I. F. Gorodnitsky and B. D. Rao, “Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm,” IEEE Transactions on signal processing, vol. 45, no. 3, pp. 600–616, 1997.

⁴ M.A.T Figueiredo, R.D. Nowak, “An EM algorithm for wavelet-based image restoration,” IEEE Transactions on Image Processing, 2003;12:906–916.

⁵ A. Beck and M. Teboulle, “A fast, iterative shrinkage-thresholding algorithm for linear inverse problems,” SIAM J. Imaging Sci., 2(1):183–202, 2009.

Dictionary Update Algorithms

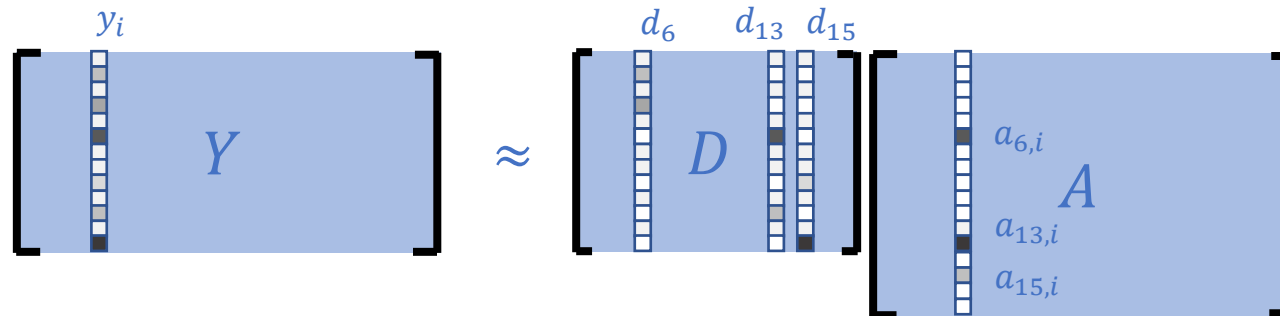
- Choices of Algorithms for ***Dictionary Update***:
 - Gradient descend
 - Least squares solution
 - Method of Optimal Directions (MOD)¹
 - K-SVD²
 - ...

¹ K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in Acoustics, Speech, and Signal Processing, 1999, Proceedings., 1999 IEEE International Conference on, vol. 5. IEEE, 1999, pp. 2443–2446.

² M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Transactions on signal processing, vol. 54, no. 11, pp. 4311–4322, 2006.

Εφαρμογές της Εκμάθησης Λεξικού

Lossy Compression



- Each data vector y_i can be represented by the values of the sparse coefficients (e.g., $a_{6,i}$, $a_{13,i}$, $a_{15,i}$) and a “bitmap” that defines their locations (6, 13 and 15) – think about DCT based compression
- The dictionary (model) must also be encoded (once for the whole data) – unlike DCT where the basis is known

Denoising

Original clean image



Noisy image, 20.1721dB



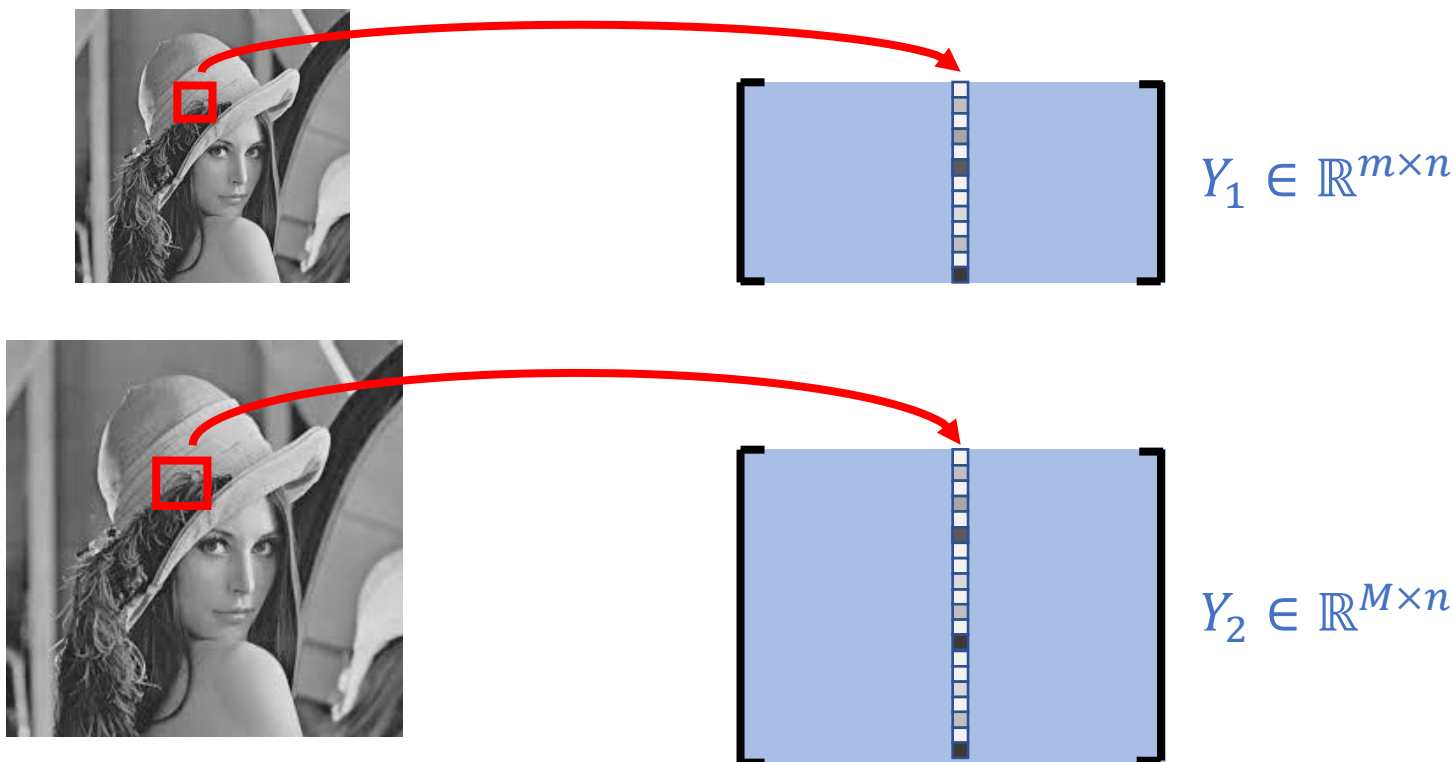
Clean Image by Global Trained dictionary, 27.5672dB



- Simply computing the sparse approximation and reconstructing has denoising effects (similar to DCT or Wavelets)
- The dictionary should not encode noise

Coupled DL – Super Resolution

- Consider we are given (the same) data of both low and hi resolution



Coupled DL – Super Resolution

- We follow the exact DL steps given data Y_1 and Y_2 so that:

$$Y_1 \approx D_1 \cdot A$$

$$Y_2 \approx D_2 \cdot A$$

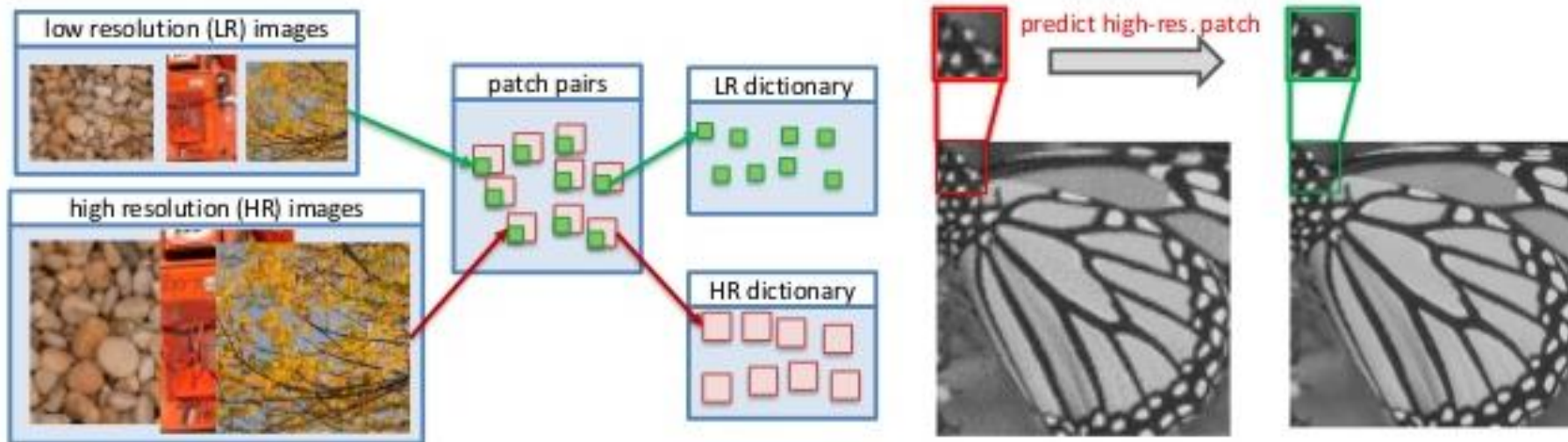
where $D_1 \in \mathbb{R}^{m \times k}$ and $D_2 \in \mathbb{R}^{M \times k}$ and $A \in \mathbb{R}^{k \times n}$ is **common**

- Thus, given some data vector $y_{1,i} \in \mathbb{R}^m$, **not included in the training/learning set**, we can compute a high-resolution version $y_{2,i} \in \mathbb{R}^M$ using the steps:
 - Sparse approximation of $y_{1,i}$ on low-resolution dictionary D_1

$$y_{1,i} = D_1 \cdot a_{1,i}$$
 - Compute high resolution as

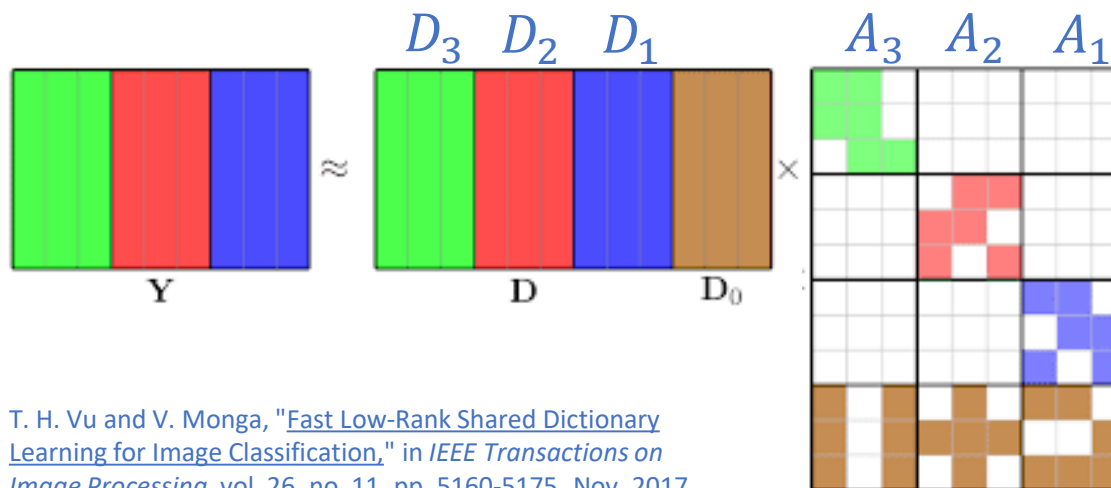
$$y_{2,i} = D_2 \cdot a_{1,i}$$

Coupled DL – Super Resolution



Discriminative DL - Classification

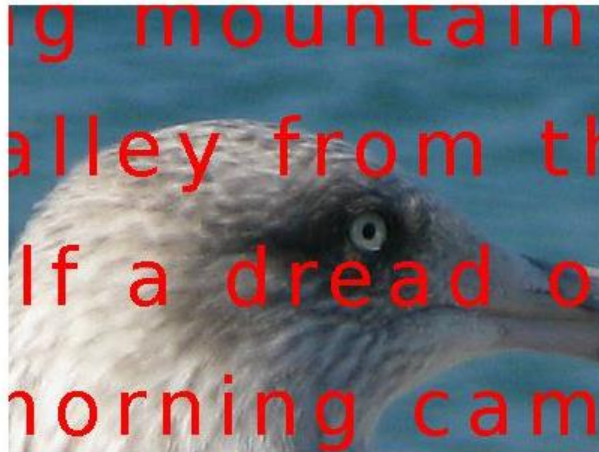
- Consider we are given a data matrix $Y \in \mathbb{R}^{m \times n}$ and labels $l \in \mathbb{R}^{1 \times n}$ where $l_i \in \{1, 2, \dots, C\}$
- The scope of discriminative DL is to learn class specific sub-dictionaries, for example:



T. H. Vu and V. Monga, "Fast Low-Rank Shared Dictionary Learning for Image Classification," in *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5160-5175, Nov. 2017.

Image Inpainting

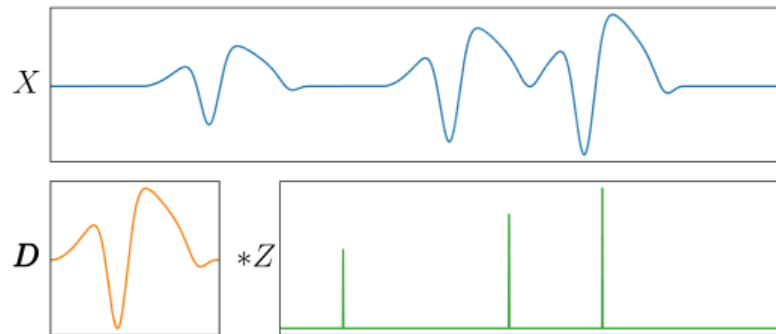
- Developments for tackling non-uniform noise, can be used to also perform image inpainting:



J. Mairal, M. Elad and G. Sapiro, "[Sparse Representation for Color Image Restoration](#)," in IEEE Transactions on Image Processing, vol. 17, no. 1, pp. 53-69, Jan. 2008.

Διάφορες άλλες μορφές της ΕΛ

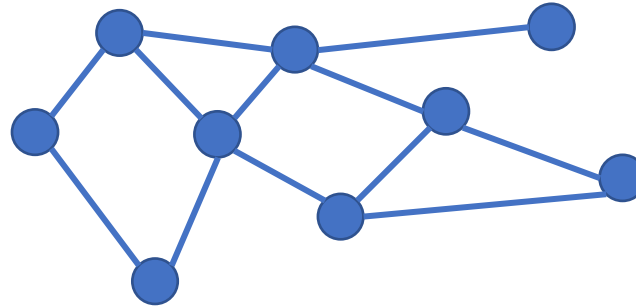
Convolutional DL



- In many application, shift invariance is required
- The convolutional DL model analyzes a signal as a sum of convolutions, where each convolution is produced by a filter-atom and a sparse signal

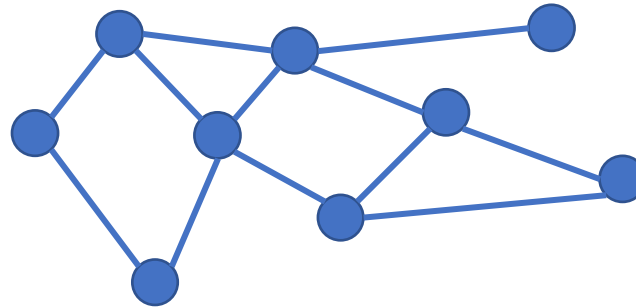
[Moreau, Thomas, and Alexandre Gramfort. "Distributed Convolutional Dictionary Learning \(DiCoDiLe\): Pattern Discovery in Large Images and Signals." arXiv preprint arXiv:1901.09235 \(2019\).](#)

Distributed DL



- Each node v has its local data matrix Y_v (camera network)
- The scope is to device a distributed algorithm to make the nodes agree (consensus) on a common dictionary
- Various approaches:
 - Diffusion type (Adapt, Combine)
 - ADMM based

Online + Distributed DL



- The data at each node is not available a-priori
- At each time instant, a new datum becomes available, which is sparsely encoded locally
- In the sequel, the nodes cooperate to perform dictionary update

Ερωτήσεις

