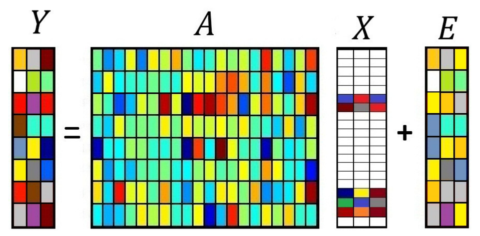
Signal Recovery with Unknown Sparsity Pattern via Multiple Measurement Vectors

Mohammad Shekaramiz, Todd K. Moon, and Jacob H. Gunther
Information Dynamics Laboratory, ECE Dept., Utah State University

Introduction

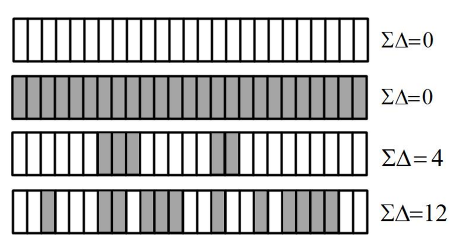


Objective:

- Recovering sparse signal X from a small set of linear noisy measurements using multiple measurement vectors (MMVs)
- **Assumption:**
  - *Sparse Clustered Pattern:* Non-zero elements of the underlying signal may appear in clusters with an unknown structure on each column of X
  - *Joint-Sparsity:* non-zero elements of X appear at the same rows (support set of the solution is the same for all columns of X)
- **Model:** $Y = A(\mathbf{s} \circ X) + E$
  $Y \in \mathbb{R}^{M\times N}$, $A \in \mathbb{R}^{M\times P}$, $\mathbf{s} \in \{0,1\}^{P\times 1}$, $X \in \mathbb{R}^{P\times N}$, $E \in \mathbb{R}^{M\times N}$, and $(M \ll P)$,
  - $\mathbf{s}$ accounts for the supports of the solution and $\circ$ denotes Hadamard product
- **Proposed algorithm:**
  - *C-SBL:* sparse Bayesian learning model for sparse signals with unknown clustered pattern

Proposed Statistical Model and Defining Priors

- Measure of clumpiness: $(\Sigma\Delta)_{\mathbf{s}} = \sum_{i=2}^{P} |s_i - s_{i-1}|$, where $\mathbf{s}$ is the support learning vector of the solution
  - There exist few transitions for the case where the supports of the solution have a block-sparsity structure
  - For example, a constant vector (all ones or all zeros) has a $\Sigma\Delta$ of 0
  - More examples:



- *Prior on the support-learning component* $\mathbf{s}$*:*
  - we model the elements of $\mathbf{s}$ as Bernoulli random variables
  $$(s_p; \omega_{0,p}, \omega_{1,p}) \sim \text{Bernoulli}(\frac{\omega_{1,p}}{\omega_{0,p} + \omega_{1,p}}), \forall p = 1, 2, ..., P$$
  $$\omega_{k,p} = e^{-\alpha(\Sigma\Delta)_{k,p}}\text{Binomial}(\Sigma_{k,p}, P, \gamma_p),\ \forall k = 0, 1$$
  - The terms $(\Sigma\Delta)_{k,p}$ and $\Sigma_{k,p}$ denote the $\Sigma\Delta$ value and the sum over all the elements of $\mathbf{s}$ for the case where $s_p = k$, respectively
  $$s_p \sim \text{Bernoulli}(\gamma_p),\ \gamma_p \sim \text{Beta}(\alpha_0, \beta_0), p = 1, \dots, P.$$
  - Initial setting: $\alpha_0 = \frac{10}{P}$ and $\beta_0 = 1 - \alpha_0$, to encourage sparsity
- *Prior on the solution-value matrix X:*
  - The columns of the solution-value matrix $X = [\mathbf{x}_1, \dots, \mathbf{x}_N]$ are assumed to be drawn i.i.d. from the following normal-gamma distribution
  $$\mathbf{x}_n \sim \mathcal{N}(0, \tau^{-1} I_P),\ \tau \sim \text{Gamma}(a_0, b_0),\ n = 1, \dots, N.$$
  - Due to the lack of prior knowledge on the entries of X, we experimentally set the hyper-parameters to $a_0 = b_0 = 10^{-3}$, endowing X *a priori* with a fairly high variance
  - a_0 and b_0 denote the shape and rate of the Gamma distribution, respectively
- *Prior on the noise component E:*
  - The entries of E are assumed to be drawn i.i.d. from a Gaussian distribution with an unknown precision ε
  $$e_{mn} \sim \mathcal{N}(0, \varepsilon^{-1}),\ m = 1, \dots, M,\ n = 1, \dots, N,$$
  $$\varepsilon \sim \text{Gamma}(\theta_0, \theta_1).$$
  - The hyper-parameters are set to $\theta_0 = \theta_1 = 10^{-3}$. This setting may vary under the required precision or prior knowledge on the noise variance

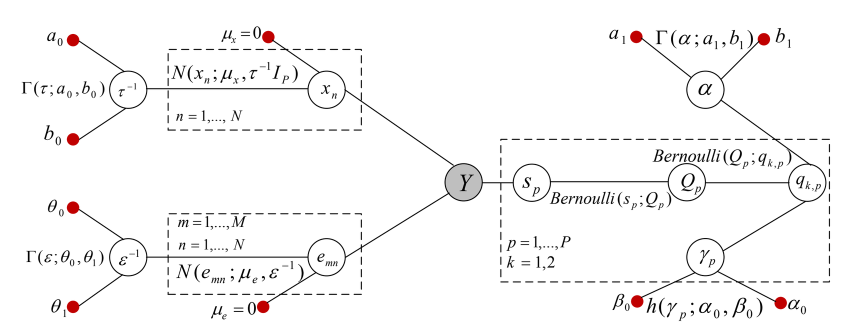
Statistical Model and Defining Priors (Continued)

- *Prior on the emphasizing factor on clumpiness of supports α:*
  $$\alpha \sim \text{Gamma}(a_1, b_1)$$
  - Initial setting: $a_1 = 5$ and $b_1 = 1$
  - The parameter $\alpha > 0$ specifies the significance of $(\Sigma\Delta)$
  - Large values of α encourage more contiguity in the support of $\mathbf{s}$, while small values of α cause $\mathbf{s}$ to have many transitions

Joint Probability Distribution of the Proposed Model

$$P(Y, \mathbf{s}, X) \propto P(Y|\mathbf{s}, X, \varepsilon)\Big(\prod_{n=1}^{N} P(\mathbf{x}_n|\boldsymbol{\mu}_x, \tau^{-1} I_P)\Big) P(\tau; a_0, b_0) P(\varepsilon; \theta_0, \theta_1)\times$$
$$\Big(\prod_{p=1}^{P}\prod_{k=0}^{1} P(s_p|\omega_{0,p}, \omega_{1,p}) P(\omega_{k,p}|\Sigma_{k,p}, \gamma_p, \alpha, \mathbf{s})\Big) P(\boldsymbol{\gamma}|\mathbf{s}, \alpha_0, \beta_0) P(\alpha|\mathbf{s}, a_1, b_1)$$

Graphical Model of the Proposed Bayesian Model



C-SBL Algorithm

C-SBL Algorithm for sparse signal recovery of either SMV or MMVs:

$\{\Theta^{(i)}\}_{i=1 \text{ to } N_{collect}} =$ **C-SBL**$(Y, A, \Theta_0, N_{burn-in}, N_{collect})$
For $Iter = 1$ to $N_{burn-in} + N_{collect}$
  % Support-learning vector component
  For $p = 1$ to P
    $\tilde{y}_{mn}^{-p} = y_{mn} - \sum_{l\neq p}^{P} a_{ml} s_l x_{ln}, \quad \forall m = 1$ to M
    $c = \frac{1-\gamma_p}{\gamma_p}\frac{\Sigma_{1,p}}{P+1-\Sigma_{1,p}} e^{-\alpha\left((\Sigma\Delta)_{0,p} - (\Sigma\Delta)_{1,p}\right)}$
    $k = \frac{\varepsilon}{2}\big((\|\mathbf{a}_p\|_2^2(\sum_{n=1}^{N} x_{pn}^2)) - 2\mathbf{a}_p^T(\sum_{n=1}^{N} x_{pn}\tilde{\mathbf{y}}_n^{-p})\big)$
    $(s_p|-) \sim \text{Bernoulli}(\frac{1}{1+ce^k})$
    % Solution-value matrix component
    For $l = 1$ to P
      $\Sigma_x = (\tau + \varepsilon s_l^2 \|\mathbf{a}_l\|_2^2)^{-1}$
      $\bar{\boldsymbol{\mu}} = \varepsilon s_l \Sigma_x \mathbf{a}_l$
      For $k = 1$ to N
        $\tilde{\mathbf{y}}_k^{-l} = \mathbf{y}_k - A(\mathbf{s} \circ \mathbf{x}_k) + s_l x_{l,k} \mathbf{a}_l$
        $\mu_x = \bar{\boldsymbol{\mu}}^T \tilde{\mathbf{y}}_k^{-l}$
        $(x_{l,k}|-) \sim \mathcal{N}(\mu_x, \Sigma_x)$
      End For
    End For
    $(\gamma_p|-) \sim \text{Beta}\big(\alpha_0 + 2(1 + \sum_{k\neq p}^{P} s_k)\ ,\ \beta_0 + 2(P - \sum_{k\neq p}^{P} s_k)\big)$
  End For
  $(\tau|-) \sim \text{Gamma}(a_0 + \frac{NP}{2}, b_0 + \frac{1}{2}\|X\|_F^2)$
  $(\varepsilon|-) \sim \text{Gamma}(\theta_0 + \frac{MN}{2}, \theta_1 + \frac{1}{2}\|Y - A(\mathbf{s} \circ X)\|_F^2)$
  $(\alpha|-) \sim \text{Gamma}\big(a_1 + \sum_{p=1}^{P}(\Sigma\Delta)_{0,p}\ ,\ b_1 + \sum_{p=1}^{P}(\Sigma\Delta)_{1,p}\big)$
  If $Iter > N_{burn-in}$
    $\Theta^{(Iter - N_{burn-in})} \leftarrow \Theta$
  End If
End For

Remark: In these descriptions, conditioning on $-$, as in $(s_p|-)$, is the inference on s_p conditioning upon all relevant variables (including the observations)

Behavior of α with respect to $(\Sigma\Delta)$

- For each element s_p we have:

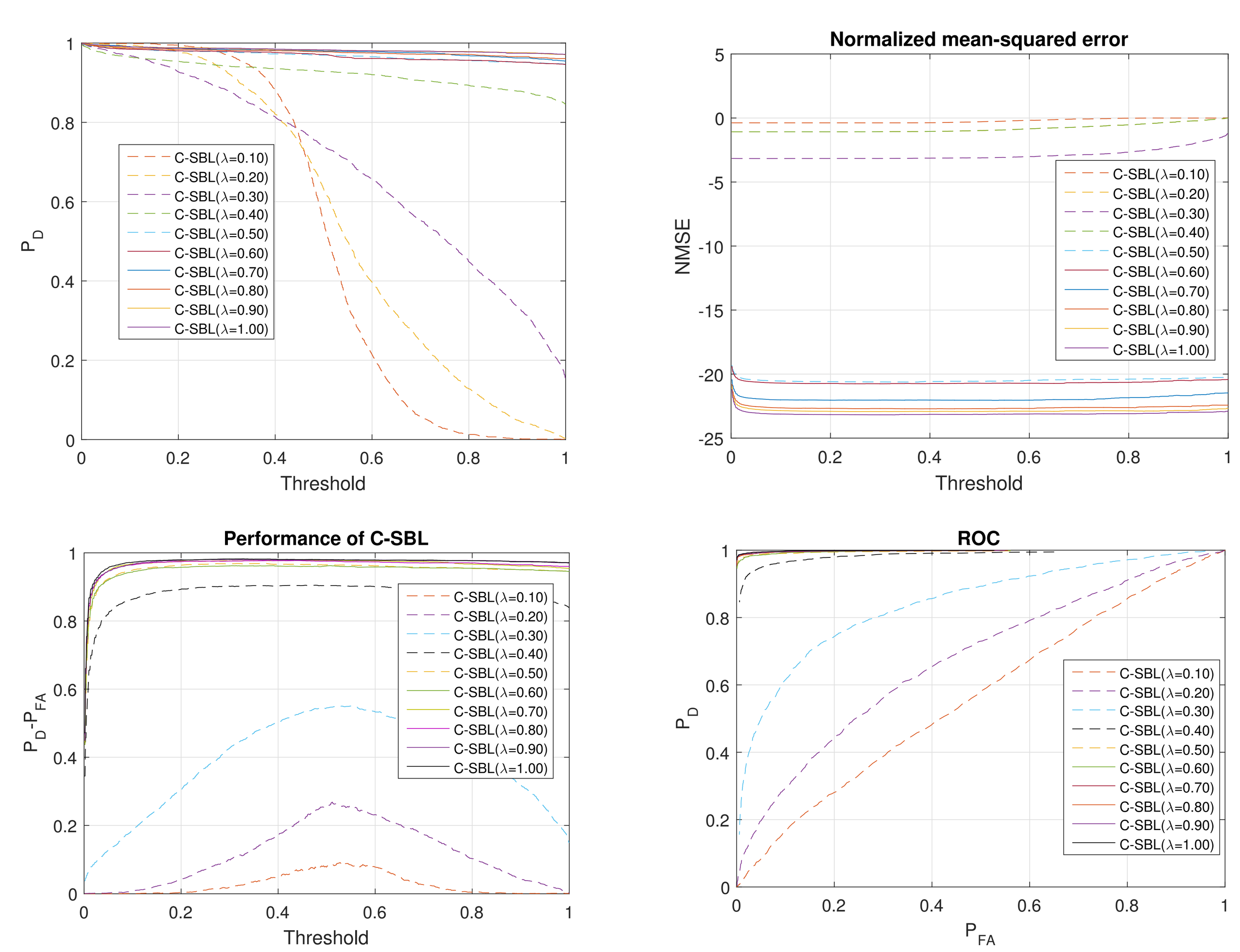
$(\Sigma\Delta)\vert_{1,p}$	$(\Sigma\Delta)\vert_{0,p}$	α
cte	*cte*	$\downarrow$
$\uparrow$	*cte*	$\downarrow$
cte	$\downarrow$	$\downarrow$
cte	$\uparrow$	$\uparrow$
$\downarrow$	*cte*	$\uparrow$

- For example consider the case where forcing either $s_p = 0$ or $s_p = 1$ does not make any change in the evaluation of $(\Sigma\Delta)$
- In this case, though it promotes the clumpiness in the solution, it discourages the solution to be sparse
- Therefore, α needs to be decreased

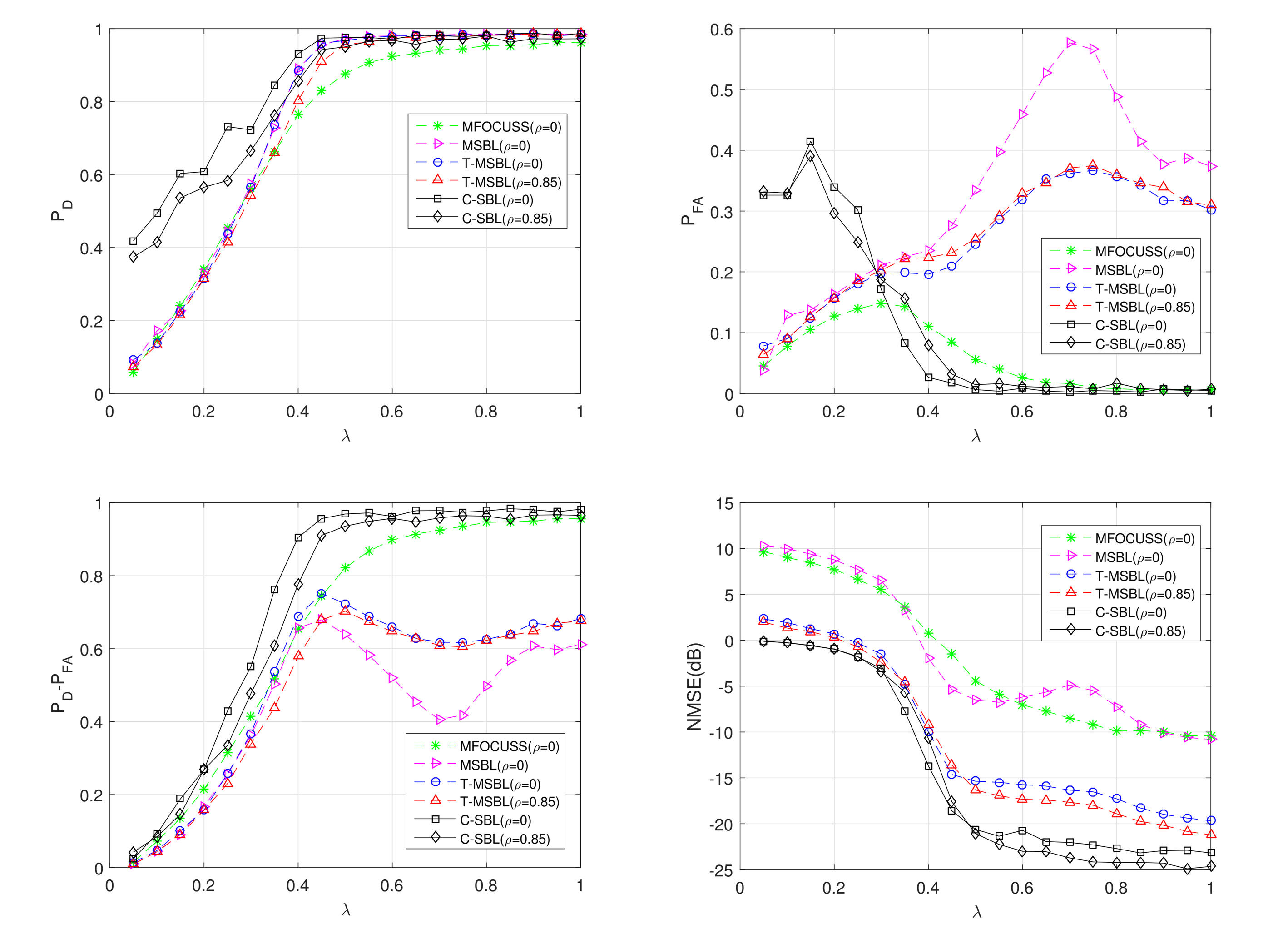
Simulation Results

- Our MMV model is a set of linear equations where
- The supports of the true solution are binary and drawn from a Bernoulli distribution in such a way to have a random clustered-sparsity structure
- The number of columns in X and Y is set to $N = 2$
- The entries of $\bar{\mathbf{x}}_n$ is drawn i.i.d. from $\mathcal{N}(\mathbf{0}, \sigma_x^2 I_P)$, where $\sigma_x^2 = 1$
- The true solution is constructed from $X = \mathbf{s} \circ \bar{X}$
- The sensing matrix $A \in \mathbb{R}^{M\times P}$, with $a_{mn} \sim \mathcal{N}(0, \frac{1}{\sqrt{M}})$, where M varies and $P = 100$
- The entries of the noise vector are drawn i.i.d. from $\mathcal{N}(0, \sigma_n^2)$, in such a way to have $SNR = 25dB$
- The measurement matrix Y is computed from $Y = AX + E$
- In all the simulations, the sparsity level is set to $K_{sp} = 25$
- In the simulations for C-SBL we set $N_{\text{burn-in}} = 500$ and $N_{\text{collect}} = 500$
- We consider two case scenarios
  - case 1: the columns of true solution X are uncorrelated i.e., $\rho = 0$
  - case 2: the columns of true solution X have correlation factor of $\rho = 0.85$
- In order to investigate the performance, we generate 200 random cases using the above settings and then averaging over all the obtained results
- In the figures, λ is the sampling rate and is defined as $\lambda := M/P$

Simulation Results (Performance of C-SBL)

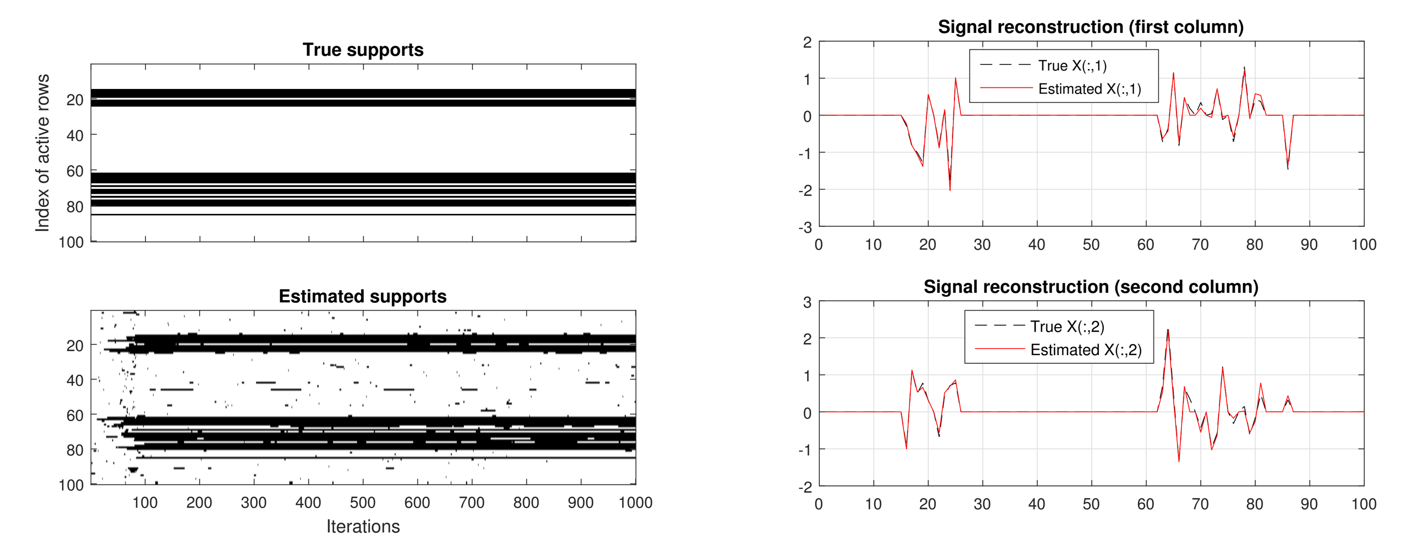


Simulation Results (Comparison with other algorithms)



A Case Scenario (Synthetic data)

- In this example $\lambda = 0.7$, $N = 2$, and $SNR = 25dB$



Conclusions

- A new algorithm for the recovery of sparse signals with unknown clustered pattern is proposed (C-SBL algorithm)
- The proposed algorithm can be used for either single- or multiple-measurement vectors in the compressive sensing (CS) applications
- Based on the simulation results, C-SBL outperforms the famous M-SBL [2], T-SBL [3], and MFOCUSS [4] algorithms

Main References

1. M. Shekaramiz, T. K. Moon, and J. H. Gunther, *Hierarchical Bayesian approach for jointly-sparse solution of multiple measurement vectors*, 48th Asilomar Conf. of Signals, Systems, and Computers, pp. 1962–1966, Nov. 2014.
2. D.P. Wipf and B. D. Rao, *An empirical Bayesian strategy for solving the simultaneous sparse approximation problem*, IEEE Trans. on Signal Processing, vol. 55, no. 7, pp. 3704–3716, 2007.
3. Z. Zhang and B. D. Rao, *Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning*, IEEE Journal of Special Topics in Signal Processing, vol. 5, no. 5, pp. 912–926, 2011.
4. S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, *Sparse solutions to linear inverse problems with multiple measurement vectors*, IEEE Transactions on Signal Processing, vol. 53, no. 7, pp. 2477–2488, 2005.