Рандомизированные алгоритмы в управлении и обработке данных I

Олег Николаевич Граничин
Санкт-Петербургский государственный университет
Oleg_granichin@mail.ru

«III ТМШ», Усадьба Гончаровых , 15 июня 2011 г.

Что ждут от науки?

- Управление

Что ждут от науки?

- Управление
- Информация

Что ждут от науки?

- Управление
- Информация
- Оптимизация

Что ждут от науки?

- Управление
- Информация
- Оптимизация

$$x \xrightarrow{\text{information}} f \xrightarrow{\text{signal}} y \xrightarrow{\text{data}} \hat{x} \xrightarrow{\text{estimates}} \text{Control}$$

Информация, данные, знания

- *Информация* (лат. научение, сведение, оповещение) – опосредованный формами связи результат отражения изменяемого объекта изменяющимся с целью сохранения системной целостности
- *Данные* – получаемые сообщения
- *Знания* обеспечивают восстановление необходимой субъекту *информации x* по получаемым в ходе экспериментов или вычислений *данным y*

Обратные задачи

Информация, данные, знания

$$y = \Phi\, x \;\left(= \Phi(x) \right)$$

y – dim m

x – dim N

Сколько надо измерений?

$$y = \Phi x + v$$

$$\Downarrow$$

$$m > N$$

Управление через обратные связи

- Обычно в условиях реального времени для «извлечения» из данных информации имеем:
  - ограничения в ресурсах
  - недостаточное количество данных с необходимым разнообразием

Детерминированные подходы часто «проваливаются»

В теории и практике возникает много трудностей, когда мы пытаемся исследовать “сложные” системы. Во многих практических приложениях традиционные детерминированные алгоритмы не дают результатов, когда система сложная. Часто это ведет к так называемым *NP*-сложным задачам

Рандомизированные алгоритмы

- Один или несколько шагов базируются на случайном выбора одного из многих детермированных правил
- Рандомизация может оказаться мощным средством для решения ряда проблем, которые кажутся нерешаемыми детерминированными методами

Для реального времени

- Рандомизация может быть использована для обогащения данных наблюдений
- Рандомизированный сценарный подход позволяет решить задачи “эффективно с высокой вероятностью” для почти всех ограничений, которых часто может быть очень много

Обогащение данных наблюдений

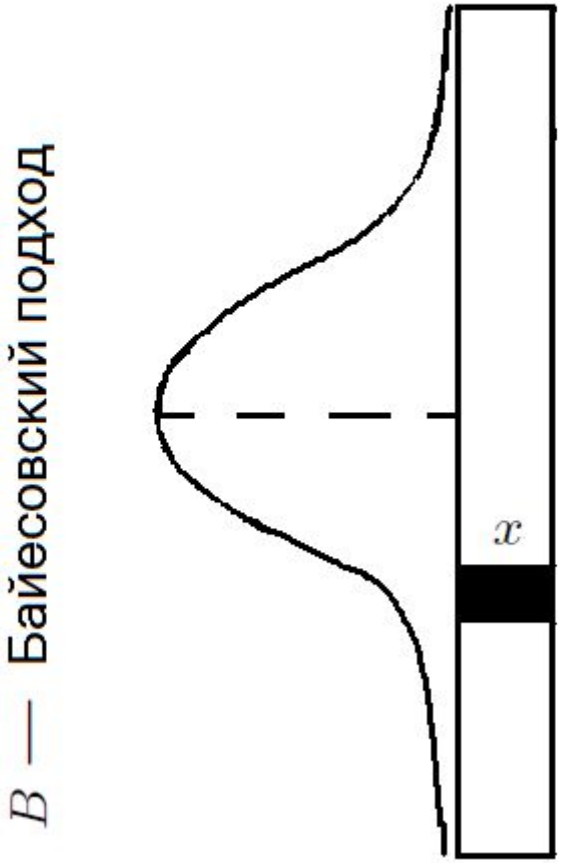
$$\text{observation} = \Phi \text{ information} + \text{noise}$$

$$y = \Phi \quad x \quad + v$$

Обогащение данных наблюдений

observation $= \Phi$ information $+$ noise

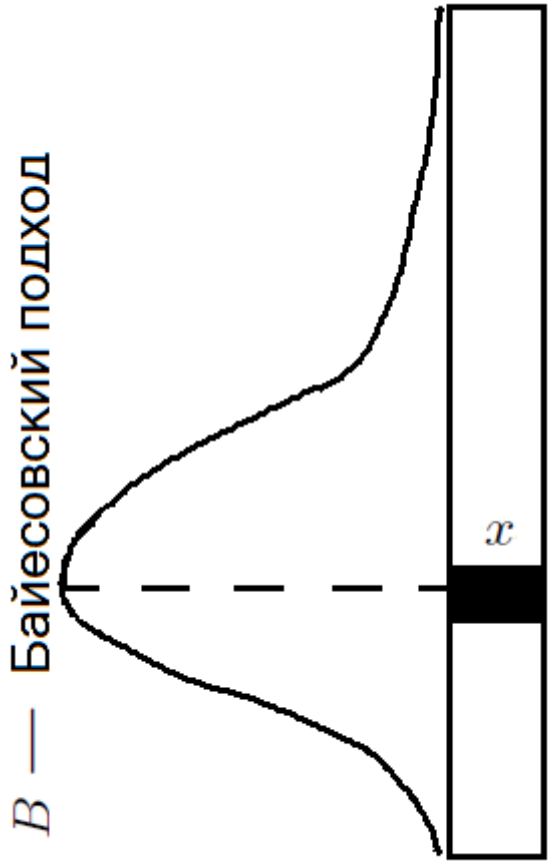
$$y = \Phi \quad x \quad + v$$



Обогащение данных наблюдений

$n = 1, 2, \ldots$

$$y_n = \Phi \quad x \quad + v_n$$

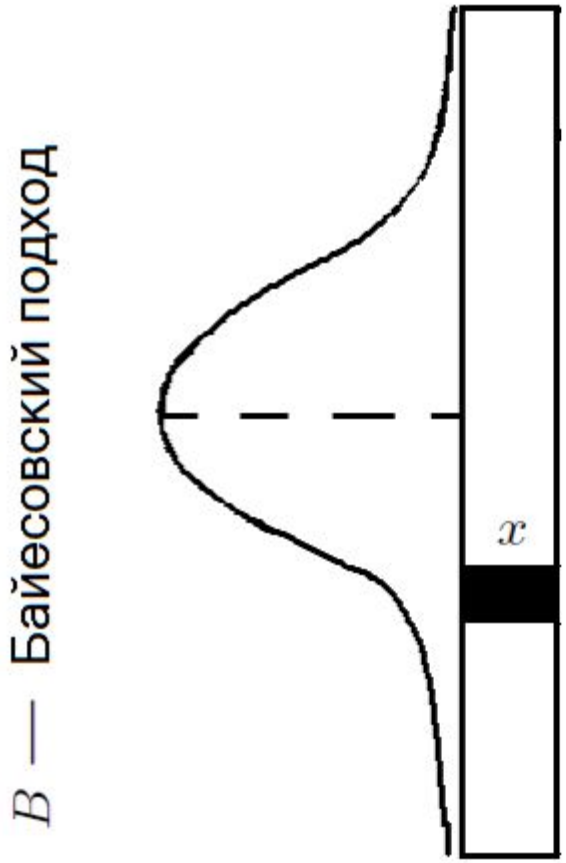


- Мы можем итеративно оценивать *B*

Обогащение данных наблюдений

$n = 1, 2, \ldots$

$$y_n = \Phi_n x_n + v_n$$

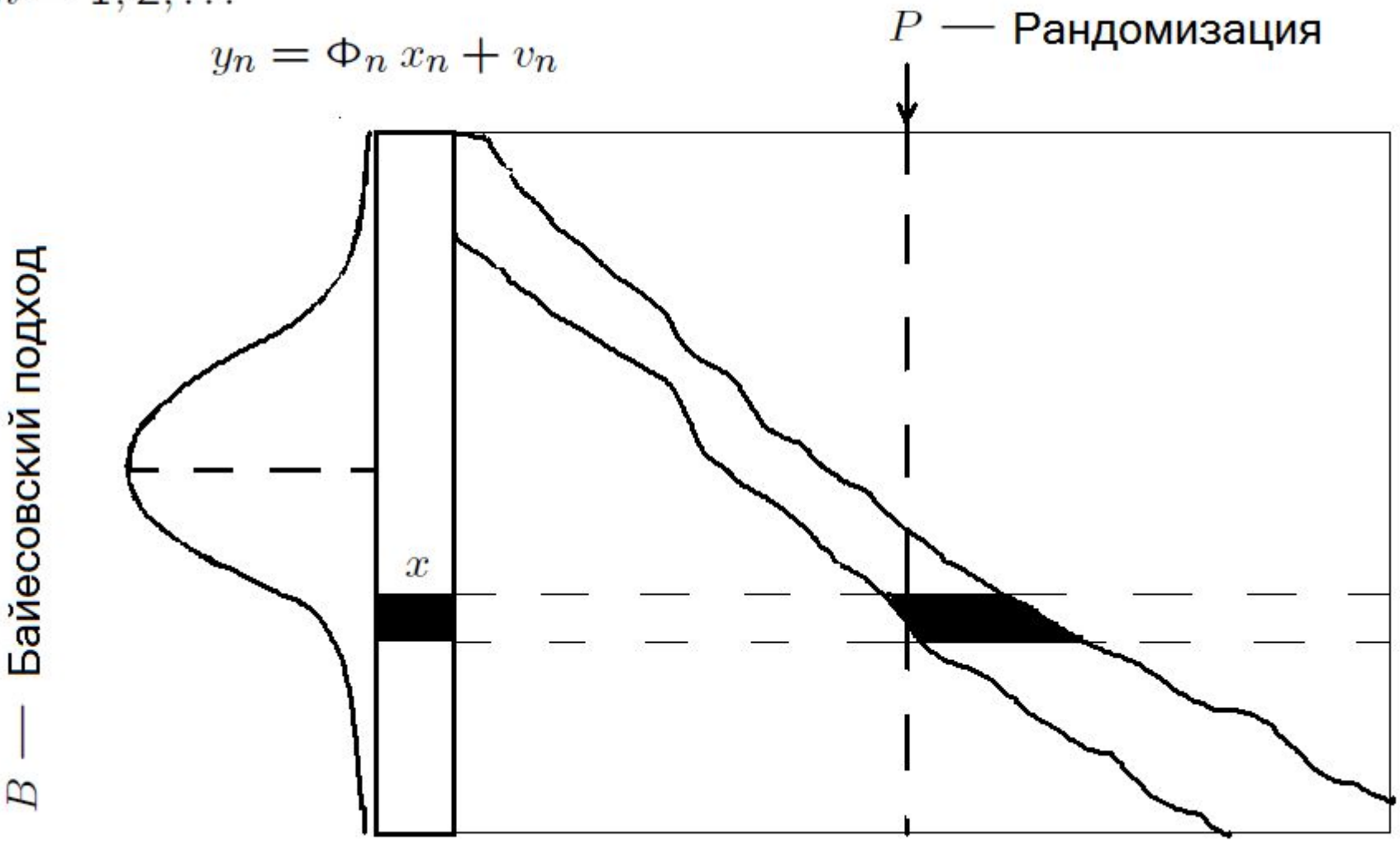


- Что можно сделать?

Обогащение данных наблюдений

$n = 1, 2, \ldots$

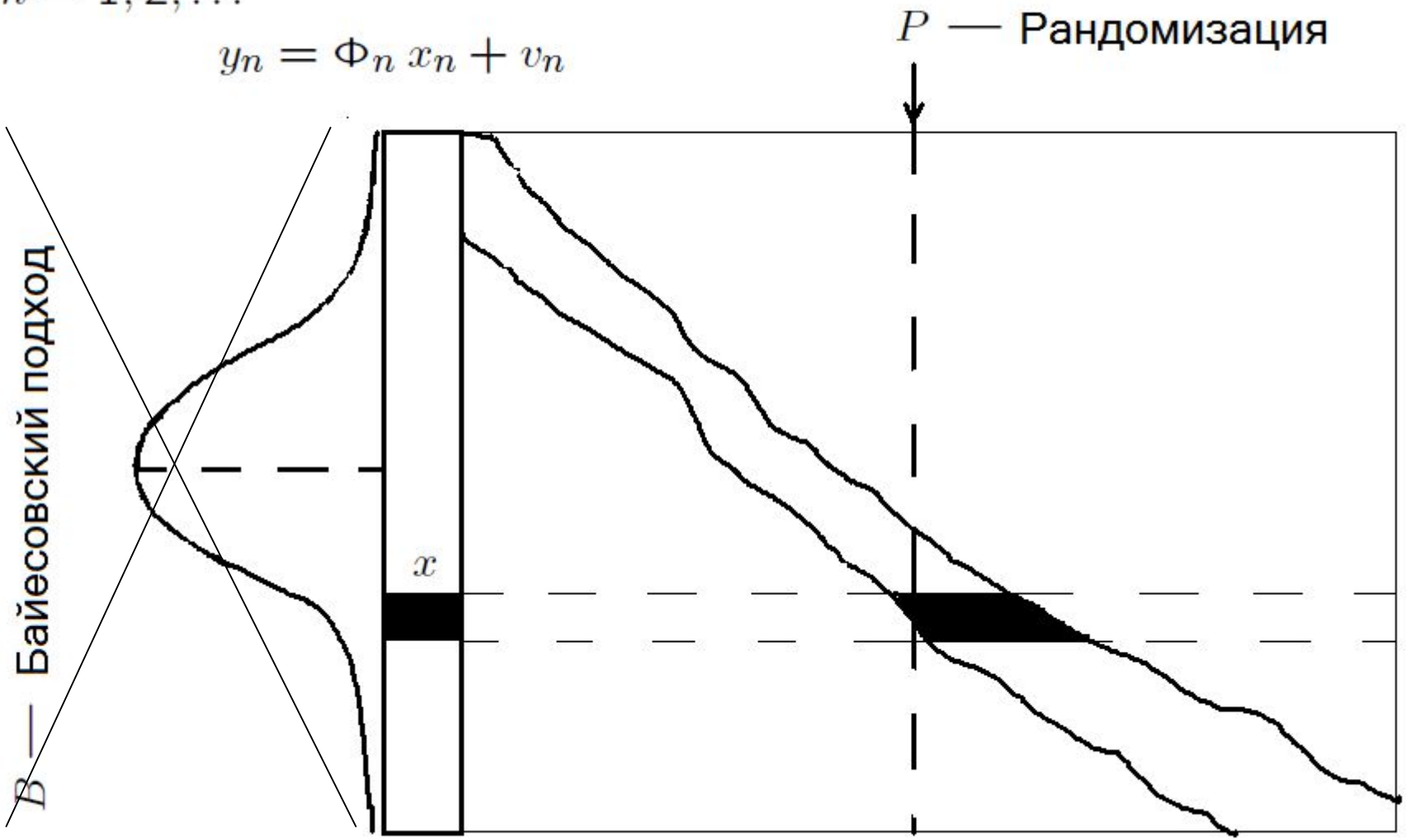
$y_n = \Phi_n x_n + v_n$



Обогащение данных наблюдений

$n = 1, 2, \ldots$

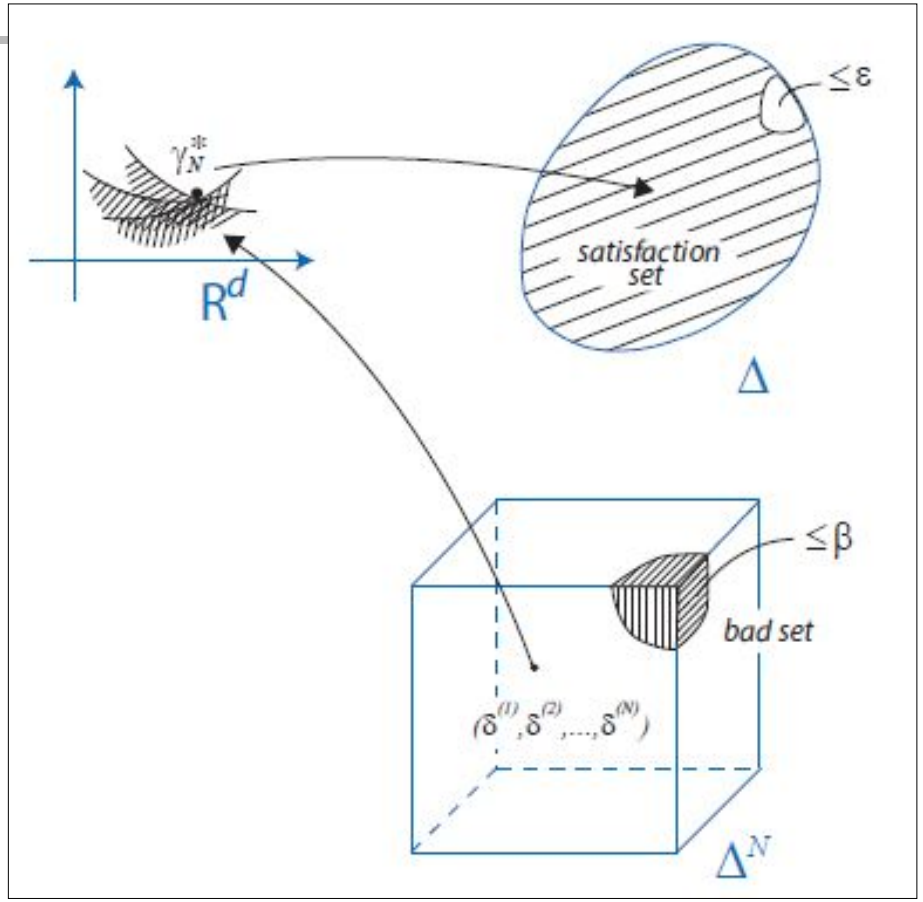
$$y_n = \Phi_n x_n + v_n$$



Сценарный подход

- Выбрать *a-priori* длину выборки *N*
- Решение задачи с *Prob=1-b*

  для почти всех условий



Рандомизация...

1930...

- Fisher (remove bias)

1950...

- Metropolis, Ulam (method Monte-Carlo)

1980-90

- Luing, Guffi, Polyak, Thibakov, Spall (fast algorithms)
- Granichin (arbitrary noise)
- Vadiyasagar (Randomized Learning Theory)

2000....

- Tempo, Campi, Calafiore, Sherbakov *etc.*

+ характеристики

- Значительное сокращение числа операций
- Снижение систематических ошибок (эффекта смещения)
- Точность обычно не зависит от размерности данных

Произвольные помехи

$$y_n = \varphi_n \theta^\star + v_n, \quad n = 1, 2, \ldots$$

$\{\varphi_n\}$ is known random with mean M_φ and $\sigma_\varphi^2 > 0$

$\{\theta^\star = 1\}$ *signal is present*

$\{\theta^\star = 0\}$ *no signal*

$\{v_n\}$ is unknown but bounded noise

Произвольные помехи

$$y_n = \varphi_n \theta^\star + v_n, \quad n = 1, 2, \ldots$$

$\{\varphi_n\}$ is known random with mean M_φ and $\sigma_\varphi^2 > 0$

$\{\theta^\star = 1\}$ *signal is present*

$\{\theta^\star = 0\}$ *no signal*

$\{v_n\}$ is unknown but bounded noise

$$(\textit{MSE}) \quad \hat{\theta}_n = \frac{\sum_{k=1}^n \varphi_k y_k}{\sum_{k=1}^n \varphi_k^2} \to \theta^\star + \frac{M_\varphi M_v}{\sigma_\varphi^2} \quad \text{a.s.}$$

$\delta = \frac{M_\varphi M_v}{\sigma_\varphi^2} + \frac{1}{2}$ desision level

If $\hat{\theta}_n < \delta$ then — otherwise +

Рандомизированный МНК

$$\Delta_n = \varphi_n - M_\varphi$$

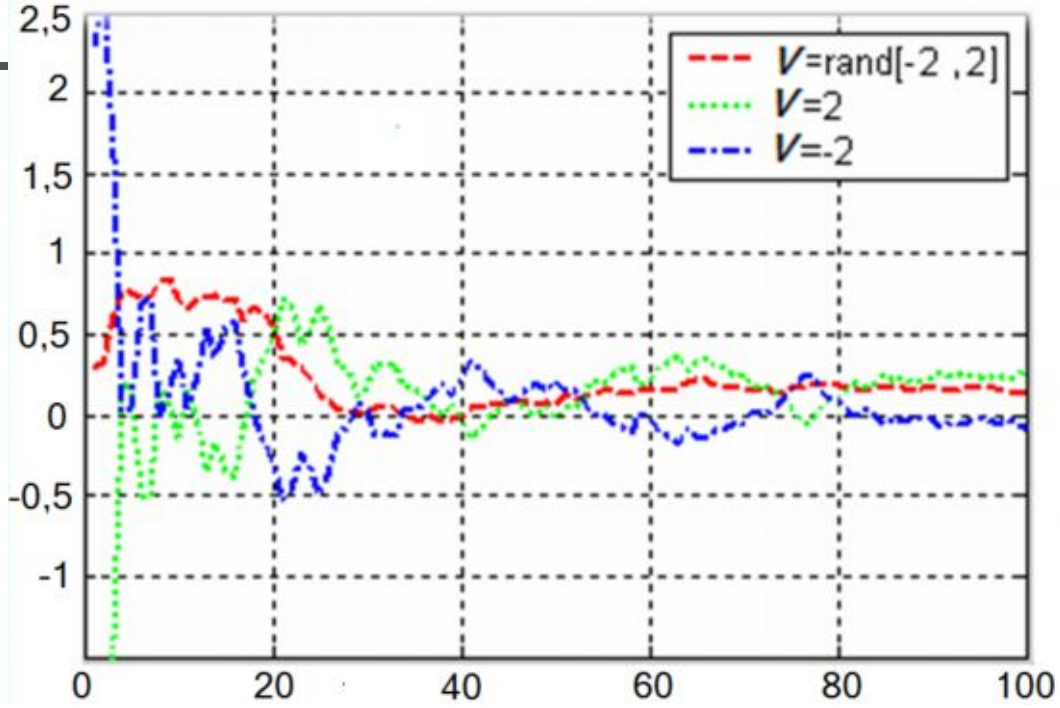
$$\Delta_n y_n = \Delta_n^2 \theta^\star + \Delta_n M_\varphi \theta^\star + \Delta_n v_n$$

$$\frac{\sum_{k=1}^n \Delta_k y_k}{\sum_{k=1}^n \Delta_k^2} = \theta^\star + \frac{\sum_{k=1}^n \Delta_k (M_\varphi \theta^\star + v_k)}{\sum_{k=1}^n \Delta_k^2}$$

$$(\ RMSE\ )\quad \hat{\theta}_n = \frac{\sum_{k=1}^n \Delta_k y_k}{\sum_{k=1}^n \Delta_k^2} \to \theta^\star \quad \text{a.s.}$$

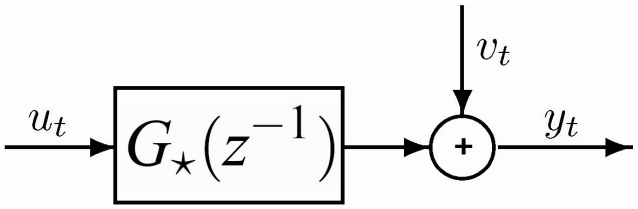
Моделирование

$$\{\varphi_n\} \text{ i.i.d.: } \varphi_n \in [0.5, 1.5], \quad |v_n| \leq 2$$



Granichin O. N., Molodtsov S. L. и др. Method of measurements with random perturbation: application in photoemission experiments // Review of scientific instruments. 79, 036103. 2008.

Конечное число наблюдений



$$u_{sn+i} = \begin{cases} \Delta_n + \bar{u}_{sn}, & \text{при } i = 0, \\ \bar{u}_{sn+i}, & \text{при } i = 1, 2, \ldots, s-1, \end{cases}$$

$n = 0, \ldots, N_\Delta - 1,$

$\bar{u}_t = \mathcal{U}_t(y_t, y_{t-1}, \ldots, \bar{u}_{t-1}, \ldots),\ t \geq 0$

1. $\hat{y}_{sn+k-1}(\theta) = \Delta_n \theta_k + \sum_{i=1}^{k-1} \theta_{k-i} \bar{u}_{sn+k-l-i},$
   $k = 1, \ldots, s.$
2. $\epsilon_t(\theta) = y_t - \hat{y}_t(\theta),\ t = 1, \ldots, N.$
3. $f_{sn+k-1}(\theta) = \Delta_n \epsilon_{sn+k-1}(\theta),\ k = 1, \ldots, s.$
4. Выберем $M > 2s$. $N_\Delta - 1$
   Вычислим $g_i^k(\theta) = \sum_{n=0} h_{i,ns+k} \cdot f_{ns+k-1}(\theta),$
   $i = 1, \ldots, M-1$
   $(h_{i,1}, \ldots, h_{i,N})$ из нулей и единиц .
5. Выберем q из $[1; M/2s]$.
   $\widehat{\Theta}^k$ не менее q из $g_i^k(\theta)$ строго $>$ нуля и
   не менее q строго меньше нуля.

$$\widehat{\Theta} = \bigcap_{k=1}^{s} \widehat{\Theta}^k.$$

$$Prob(\theta_\star \in \widehat{\Theta}) \geq 1 - 2sq/M$$

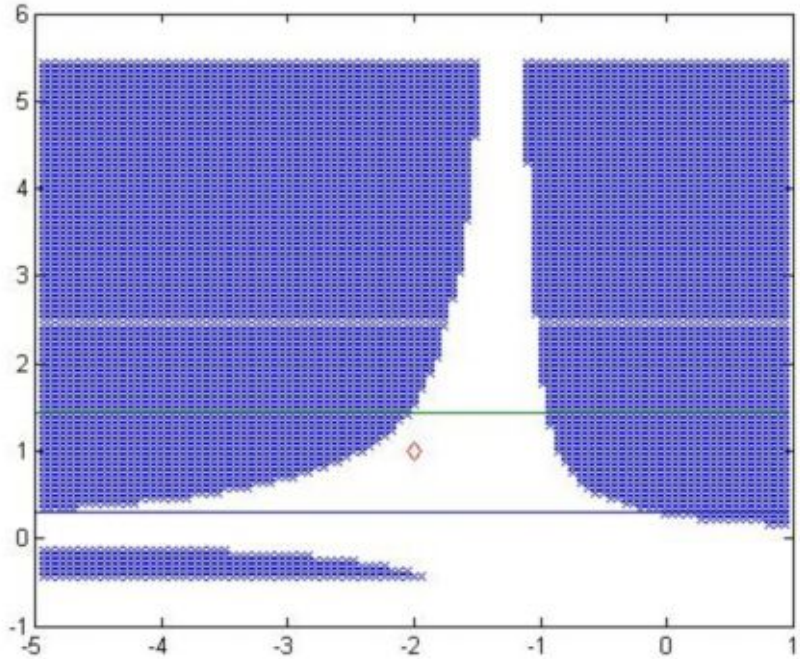
Пример

$$y_t + a_\star^1 y_{t-1} + y_{t-2} = b_\star^1 u_{t-1} + 1.6 u_{t-2} + v_t,\ t = 1, 2, \ldots, N$$

$$Prob(\theta_\star \in \widehat{\Theta}) \geq 1 - 2sq/M$$

$$95\% = 1 - 2 \cdot 2 \cdot 6/480$$

- Граничин О.Н. *Неасимптотическое доверительное множество для параметров линейного объекта управления при почти произвольных помехах* // АиТ, 2011.
- Campi M.C., Weyer E. *Non-asymptotic confidence sets for the parameters of li transfer functions* // IEEE TAC, 2010.



Многомерная оптимизация

$x_1, x_2, \ldots$ — точки наблюдений $\in \mathbb{R}^d$

$y_1, y_2, \ldots$ — наблюдения $\in \mathbb{R}^1$

$$y_n = F(x_n, w_n) + v_n$$

$w_1, w_2, \ldots$ — неконтролируемые случайные возмущения с неизвестным распределением $\mathsf{P}(\cdot)$

$v_1, v_2, \ldots$ — неизв., но ограниченные (неслучайн.)

Цель

$$f(x) = \int F(x, w)\mathsf{P}(dw) \to \min$$

Конечно-разностные аппроксимации

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \alpha_n \frac{Y_n^+ - Y_n^-}{2\beta_n},$$

$$Y_n^{\pm} = \begin{pmatrix} F\left(w_{2r(n-1)+\frac{3\pm1}{2}}, \hat{\theta}_{n-1} \pm \beta_n e_1\right) \\ F\left(w_{2r(n-1)+\frac{7\pm1}{2}}, \hat{\theta}_{n-1} \pm \beta_n e_2\right) \\ \vdots \\ F\left(w_{2rn-\frac{1\pm1}{2}}, \quad \hat{\theta}_{n-1} \pm \beta_n e_r\right) \end{pmatrix}$$

Конечно-разностные аппроксимации

- Число наблюдений = $2d$

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \alpha_n \frac{Y_n^+ - Y_n^-}{2\beta_n},$$

$$Y_n^{\pm} = \begin{pmatrix} F\left(w_{2r(n-1)+\frac{3\pm1}{2}}, \hat{\theta}_{n-1} \pm \beta_n e_1\right) \\ F\left(w_{2r(n-1)+\frac{7\pm1}{2}}, \hat{\theta}_{n-1} \pm \beta_n e_2\right) \\ \vdots \\ F\left(w_{2rn-\frac{1\pm1}{2}}, \quad \hat{\theta}_{n-1} \pm \beta_n e_r\right) \end{pmatrix}$$

Алгоритм SPSA

$\Delta_1, \Delta_2, \ldots$ — probing sampling with Bernoulli distribution (simultaneous perturbation) $\Delta_n = \begin{pmatrix} \pm 1 \\ \pm 1 \\ \vdots \\ \pm 1 \end{pmatrix} \in \mathbb{R}^d$

$$\widehat{\theta}_0 \in \mathbb{R}^d$$

$$x_n^{\pm} = \widehat{\theta}_{n-1} \pm \beta_n^{\pm} \Delta_n$$

$$\widehat{\theta}_n = \widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n^+ + \beta_n^-} \Delta_n (y_n^+ - y_n^-)$$

Алгоритм с одним измерением

$$x_n = \widehat{\theta}_{n-1} - \beta_n \Delta_n$$

$$\widehat{\theta}_n = \widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n} \Delta_n y_n$$

Алгоритм с одним измерением

$$x_n = \widehat{\theta}_{n-1} - \beta_n \Delta_n$$

$$\widehat{\theta}_n = \widehat{\theta}_{n-1} - \frac{\alpha_n}{\beta_n} \Delta_n y_n$$

- Сокращение числа наблюдений до 1 или 2 вместо $2d$

Обоснование SPSA

$$E(\theta_n - \frac{\alpha_n}{\beta_n}\Delta_n f(\theta_n + \beta_n \Delta_n)) \approx$$

$$\theta_n - \frac{\alpha_n}{\beta_n} E(\Delta_n f(\theta_n) + \frac{\beta_n \Delta_n \Delta_n \nabla f(\theta_n)}{2}) =$$

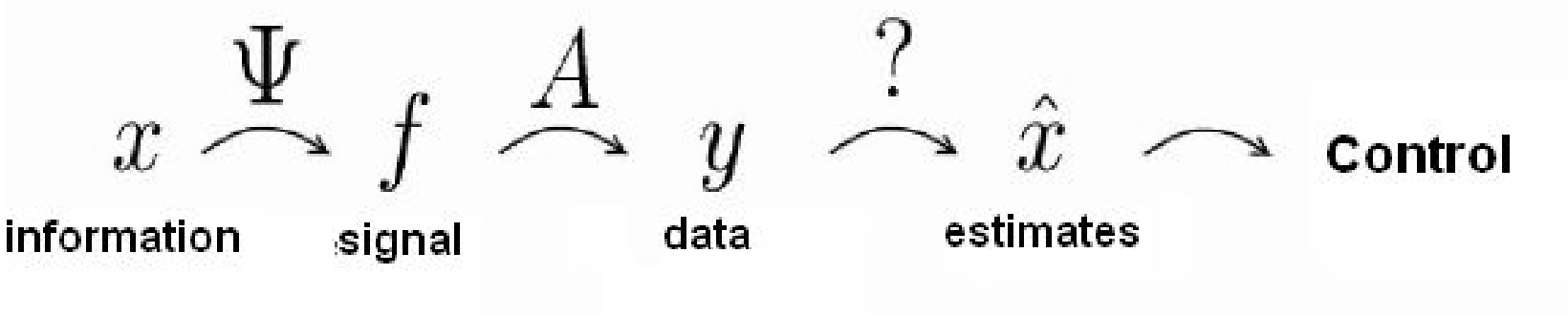
$$\theta_n - \frac{\alpha_n}{2} \nabla f(\theta_n).$$

Преимущества SPSA

- Асимптотически оптимальная скорость сходимости
- Min число измерений на итерации
- Состоятельность при почти произвольных помехах
- Работоспособность в нестационарных задачах
- Легко может быть реализован на квантовом компьютере (см. 16/06/2011)

Рандомизация получения данных и ℓ_1-оптимизация (Compressive Sensing)

Информация, получение данных, управление



$$\Phi = A\,\Psi$$

$$y = \Phi\, x$$

Теория информации

Теорема Шеннона-Найквиста-Котельникова

$$x = \{f(t) : t \in \mathbb{R}\} \in L_2(\mathbb{R})$$

$$N = 2\omega_{\max}$$

$$f(t) = \sum_{i \in \mathbb{Z}} y_i sinc(\pi(\omega_{\max} t - i),$$

$$y_i = f(\frac{i}{\omega_{\max}}),\ sinc(t) = \frac{sin(t)}{t}.$$

От 1D к 2D, 3D

1D – 1940-80

- *N=10^3*

2D – 1980-90

- *N=10^6* – Large

3D – 2000-...

- *N=10^9* – Huge

$m << N$ *Возможно ли?*

Sparse (compressive) signals

s-sparse vector

x=col(0,…,0,1,0,…,7,0,…,0,3,0,…) – 3-sparse

s-sparse signal

$f \quad N \times 1 \quad \mathbb{R}^N$

$f[n], \; n = 1, 2, \ldots, N$

$\{\psi_j\}_{j=1}^N$

$\Psi = (\psi_1, \psi_2, \ldots, \psi_N) \quad N \times N$

$$f = \sum_{j=1}^{N} x[j]\psi_j, \quad \text{or} \quad f = \Psi x,$$

$$x[j] = \langle x, \psi_j \rangle = \psi_j^T x,$$

Трансформирующее кодирование

Недостатки:

- все *N* данных нужны
- все *N* элементов должны быть рассчитаны, хотя многие из них нули
- *2s элементов должно быть закодировано*

Новая парадигма

Compressive sensing

- $m << N$
- compressive (sparse) signals

Candes E., Romberg J., Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information // IEEE Trans. Inform. Theory. Vol. **52**. No. 2. Feb. 2006. P. 489–509.

Donoho D. Compressed sensing // IEEE Trans. Inform. Theory. Vol. **52**. No. 4. Apr. 2006. P. 1289–1306.

Compressive Sensing

$$m < N$$

$$\{a_i\}_{i=1}^{m}$$

$$y_i = \langle a_i, f \rangle$$

$$y \quad m \times 1$$

$$y = Af = A\Psi x = \Phi x,$$

$$\Phi = A\Psi \qquad m \times N$$

Задачи

- Как выбрать универсальную матрицу *A?*
- Каков алгоритм восстановления сигнала?

Матрица *A*

- Restricted Isometric Property

$$\sqrt{1-\delta} \leq \frac{||\Phi z||_2}{||z||_2} \leq \sqrt{1+\delta}$$

- Incoherence

$$\mu(A,\Psi) = \sqrt{N} \max_{i,j} \frac{|\langle a_i, \psi_j \rangle|}{||a_i||_2}$$

Универсальная случайная матрица *A*

- Gaussian *N(0,1/m)*
- Bernoulli $\mathrm{P}(a[i,j] = \pm 1/\sqrt{m}) = \frac{1}{2}$

- или $a[i,j] \in \{0, \pm\sqrt{3/m}\}$

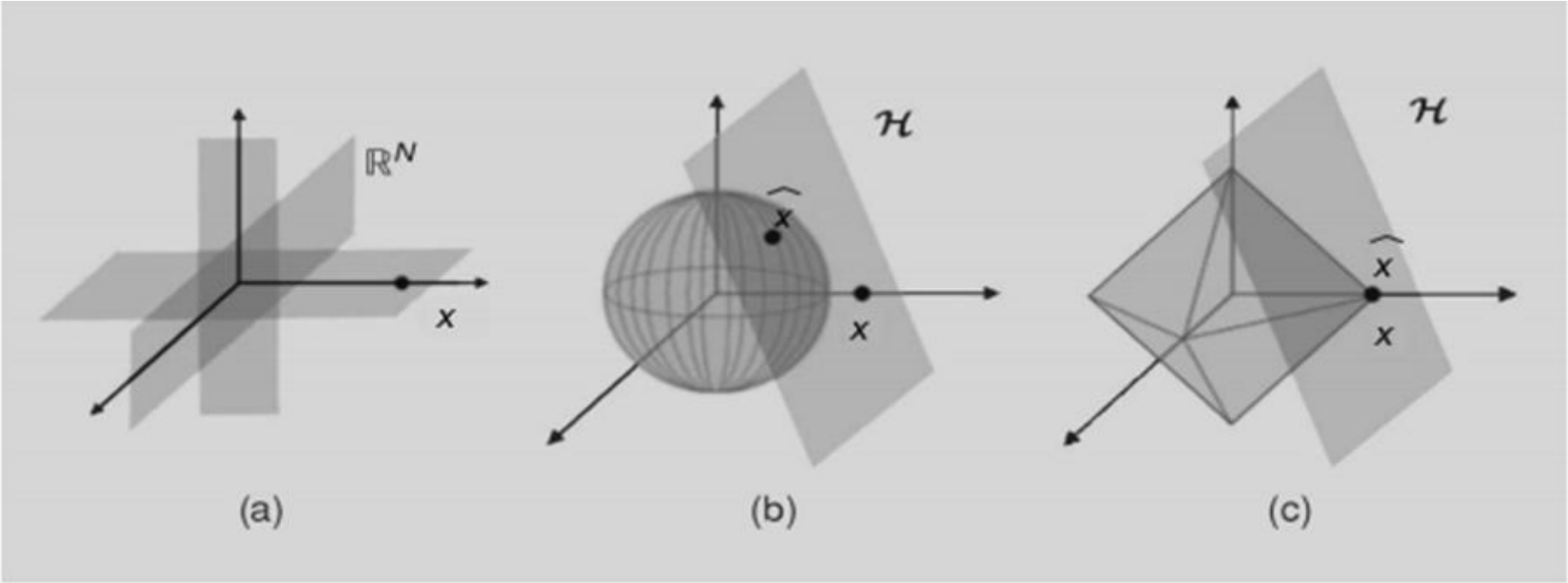
$$m \geq c\, s \log(N/s)$$

Алгоритм восстановления сигнала

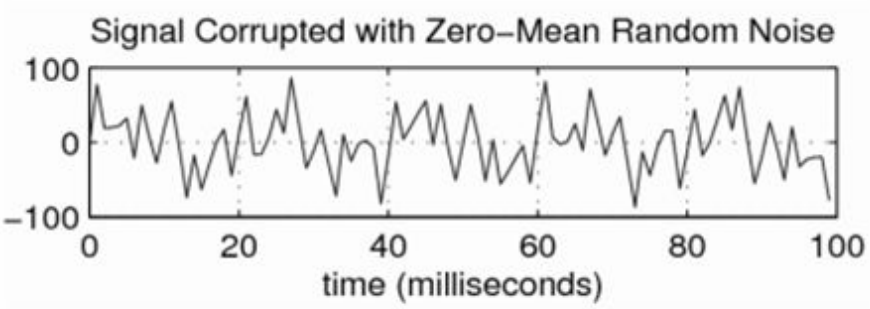
Если выполнено условие
RIP (2s) или RIP(3s), то достаточно
решить задачу ℓ_1-оптимизации

$$||x||_1 = \sum_j |x[j]| \to \min \ : \quad y = \Phi x$$

ℓ_1-оптимизация



Пример

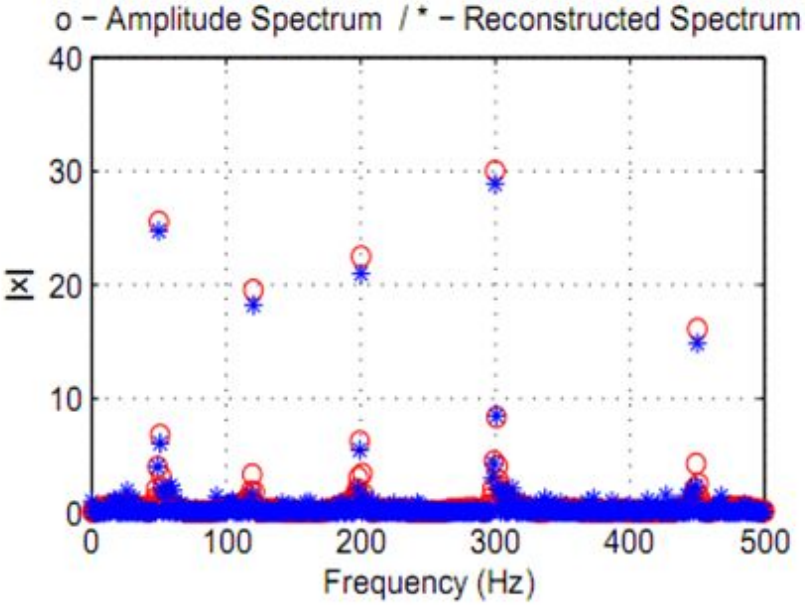


$$f(t) = \sum_{j=0}^{N} x[j] \exp^{i2\pi jt/N},$$

$$t = 0, \ldots, N-1,$$

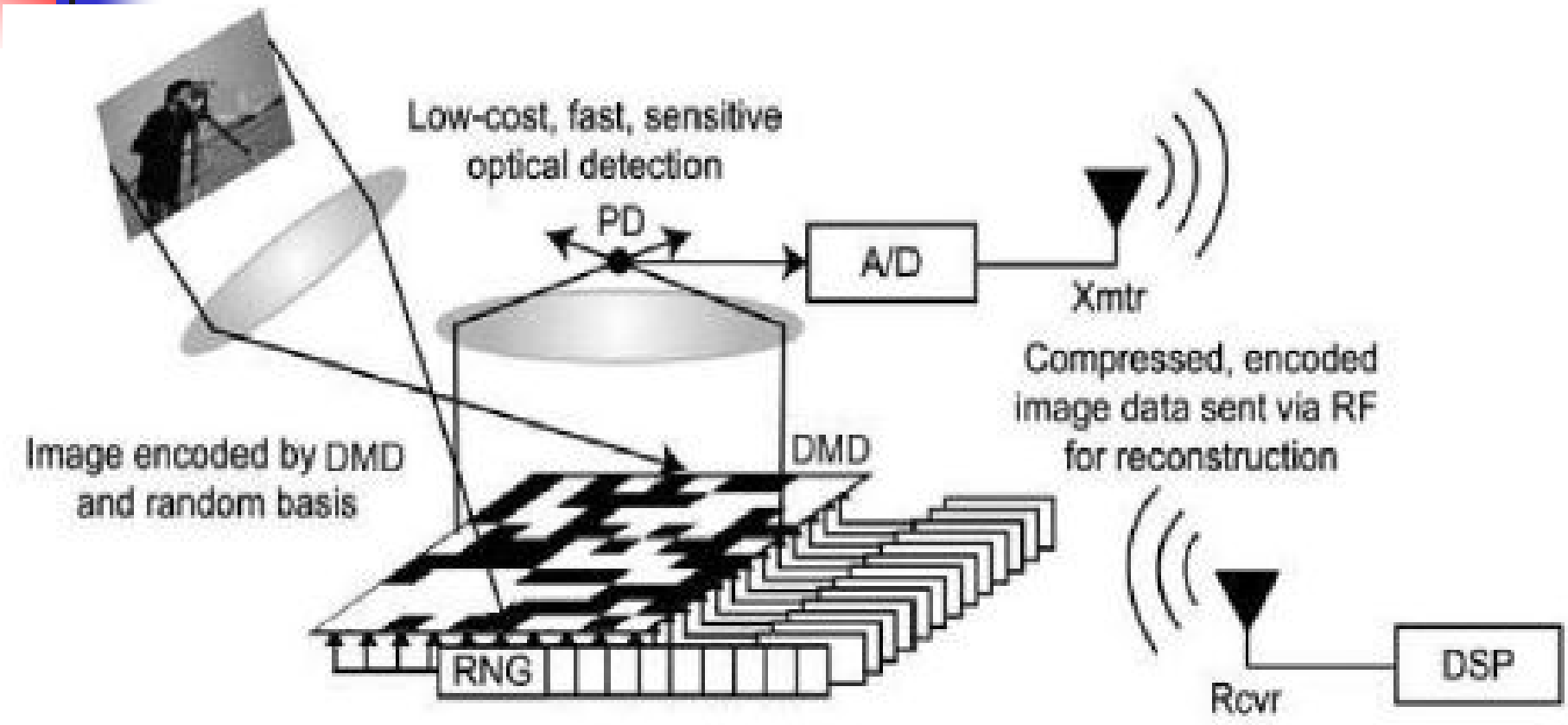
$$s << N$$

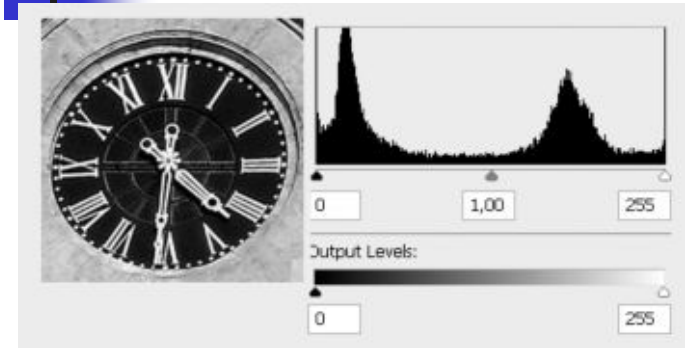
$$s = 5$$



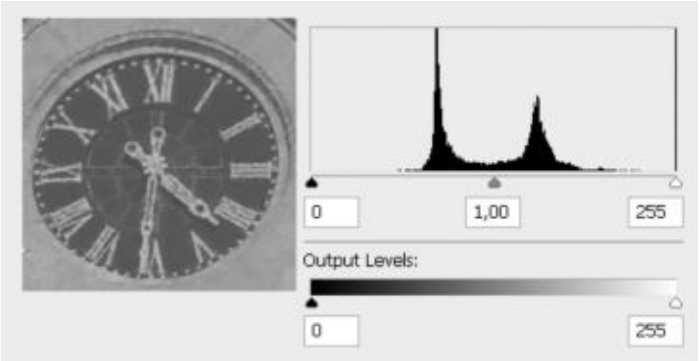
Определение частот и амплитуд 5-редкого сигнала.

Однопиксельная камера

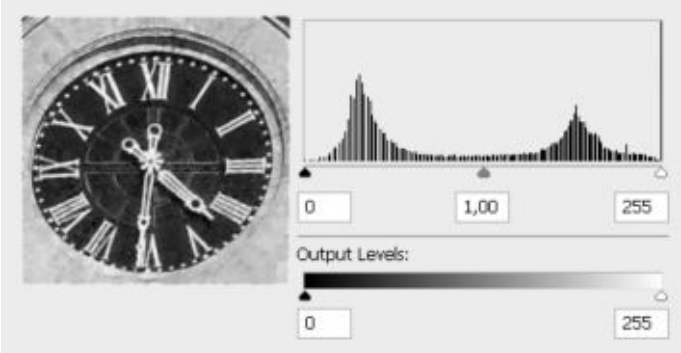



0
1,00
255
Output Levels:
0
255

Оригинал


0
1,00
255
Output Levels:
0
255

40%


0
1,00
255
Output Levels:
0
255

Другие алгоритмы восстановления

- ℓ_1-регуляризованный МНК (LASSO)
- аппроксимация ℓ_0-нормы
- минимизация невыпуклых регуляризаторов
- итеративные алгоритмы
- etc (АиТ, 2010, № 10)

Approach	Geometric/ /Combinatorial	Number of Measurements m	Number of Iterations	Worst Case Time Complexity	k-term Approximation	Noise Resilience
$\ell 1$ Embedding of Expander Codes	Combinatorial	$O\left(k(\log n)^{\log n}\right)$	-	$O(n^3)$	Compressible	No
Unique Neighborhood	Combinatorial	$O(n)$	O(n)	$O(n)$	Almost k-sparse	Sparse Noise
Unique Neighborhood	Combinatorial	$O(k \log n)$	$O(k \log n)$	$O(n \log n)$	Almost k-sparse	Sparse Noise
RIP-1	Geometric	$O\left(k \log\left(\frac{n}{k}\right)\right)$	-	$O(n^3)$	Compressible	Sparse Noise
Hash Functions	Combinatorial	$O(k \log \frac{n}{k} \log n)$	$O(k \log \frac{n}{k} \log n)$	$O(k \log^3 n)$	Compressible	Sparse Noise
All Unique Neighborhoods	Combinatorial	$O\left(k \log\left(\frac{n}{k}\right)\right)$	$O(n)$	$O\left(n \log\left(\frac{n}{k}\right)\right)$	Compressible	Sparse Noise
RIP-1	Geometric	$O\left(k \log\left(\frac{n}{k}\right)\right)$	$R = O\left(\log\left(\frac{d\|x\|_1}{\|e\|_1}\right)\right)$	$O\left(n \log\left(\frac{n}{k}\right) \log R\right)$	Compressible	Sparse Noise
Unique Neighborhood	Combinatorial	$O\left(k \log\left(\frac{n}{k}\right)\right)$	$O(k)$	$O\left(n \log\left(\frac{n}{k}\right)\right)$	Almost k-sparse	Sparse Noise

Обучение, классификация

Задача классификации заключается в отнесении некоторого входного вектора f к одному из нескольких заранее известных классов d. Задача обучения состоит в формировании правила классификации по заданной обучающей выборке пар (f_k, d_k). При этом для близкого к f_k входа классификатор должен дать близкий к d_k выходной вектор.

Calderbank R., Jafarpour S., Robert Schapire R. Compressed Learning: Universal Sparse Dimensionality Reduction and Learning in the Measurement Domain. Preprint. 2009 [Online]. Available: http://dsp.rice.edu/files/cs/cl.pdf

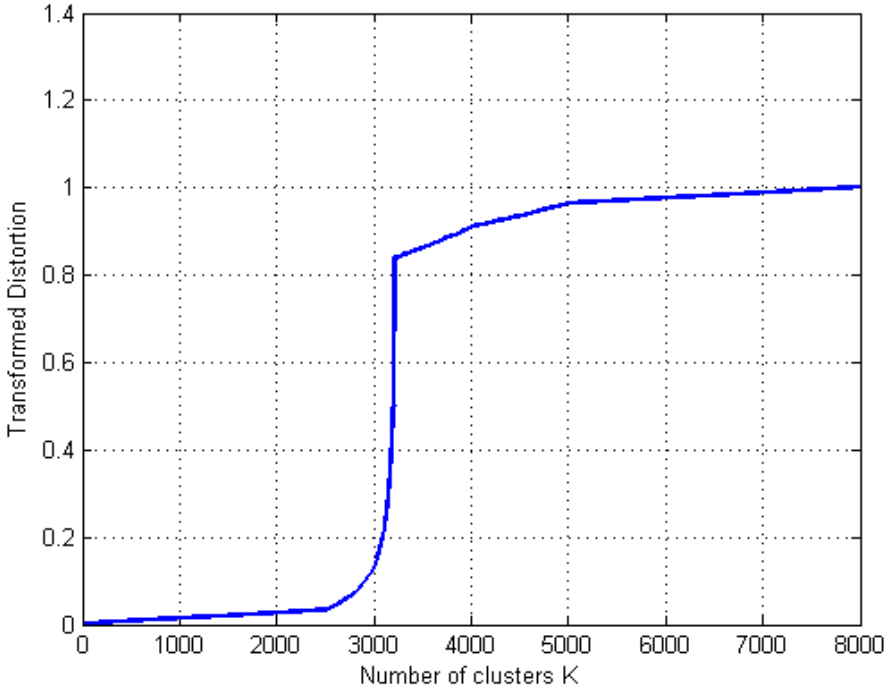
Стоит отметить важный отличительный момент такого варианта использования принципов Compressive Sensing: сжатый вектор **не требует восстановления**, его обработка происходит сразу в сжатой форме.

Устойчивая кластеризация

Рандомизированный алгоритм нахождения количества кластеров // АиТ, 2011, № 4.

Одна из сложных и важных задач кластерного анализа – определить количество кластеров в наборе данных.

Для ее решения часто используется (наряду с другими) подход, основанный на анализе изменений «Кривой искажений», которая обычно имеет «скачок» в точке, соответствующей истинному количеству кластеров



Сценарный подход

1. Выбрать параметр надежности $\beta \in (0,1)$.
2. Выбрать параметр M

$$p_m(x) = \cos(m \arccos x), \quad m = 0, 1, 2, \ldots, M.$$

3. Выбрать $N \geq M$ и $T > 1$:

$$T = \left[\frac{1}{N}\left(\frac{4C}{\beta B N} - 1\right)\right] + 1.$$

4. Выбрать T групп по N точек из $(0;1)$:

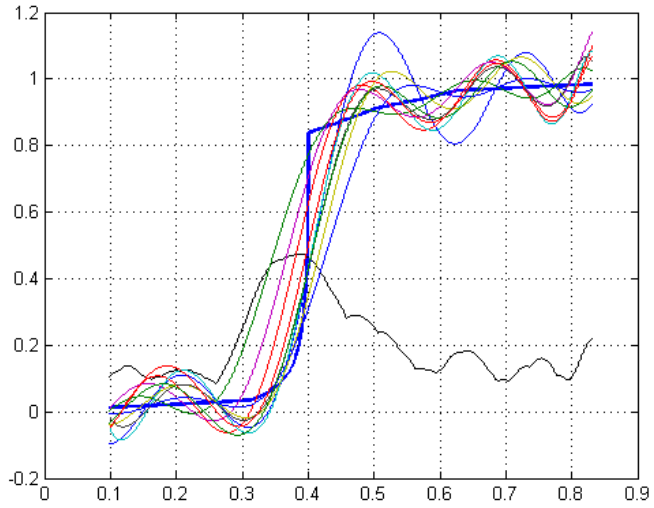
$$Z_t = \{z_{tn},\ n = 1, \ldots, N\}, \quad Z = \bigcup Z_t.$$

6. $\chi(x) = \max\limits_{t=1,\ldots,T} g_t(x) - \min\limits_{t=1,\ldots,T} g_t(x), \ x \in (0;1)$

7. Вычислить $\gamma = \max\limits_t \gamma_t$ и определить линию уровня принятия решения

$$L(x) = \frac{3B}{4} - \frac{B}{4C} h(x) - 2\gamma,$$

где $h(x) = \max\limits_{z \in [z_l(x), z_r(x)]} \max\limits_{t=1,\ldots,T} |g_t'(z)|$.

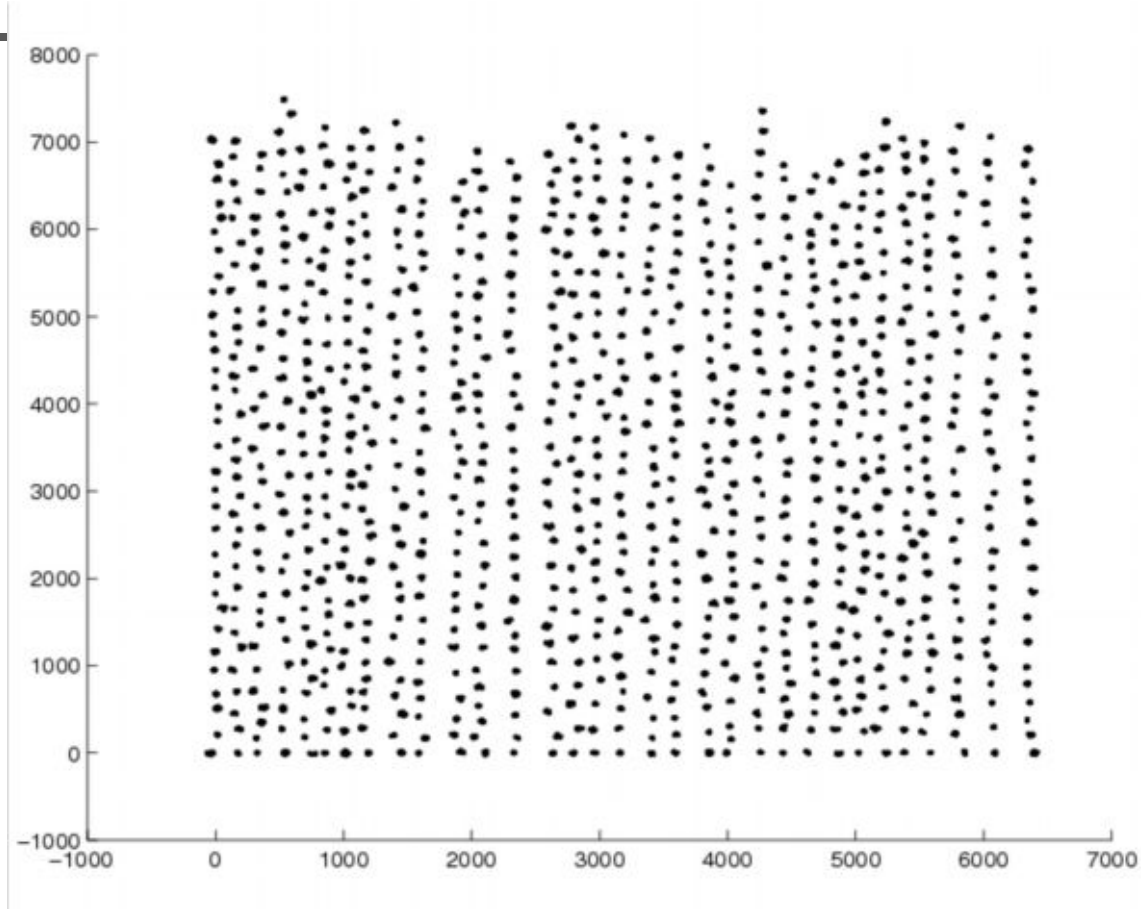


С вероятностью $P = (1-\beta)$ интервал

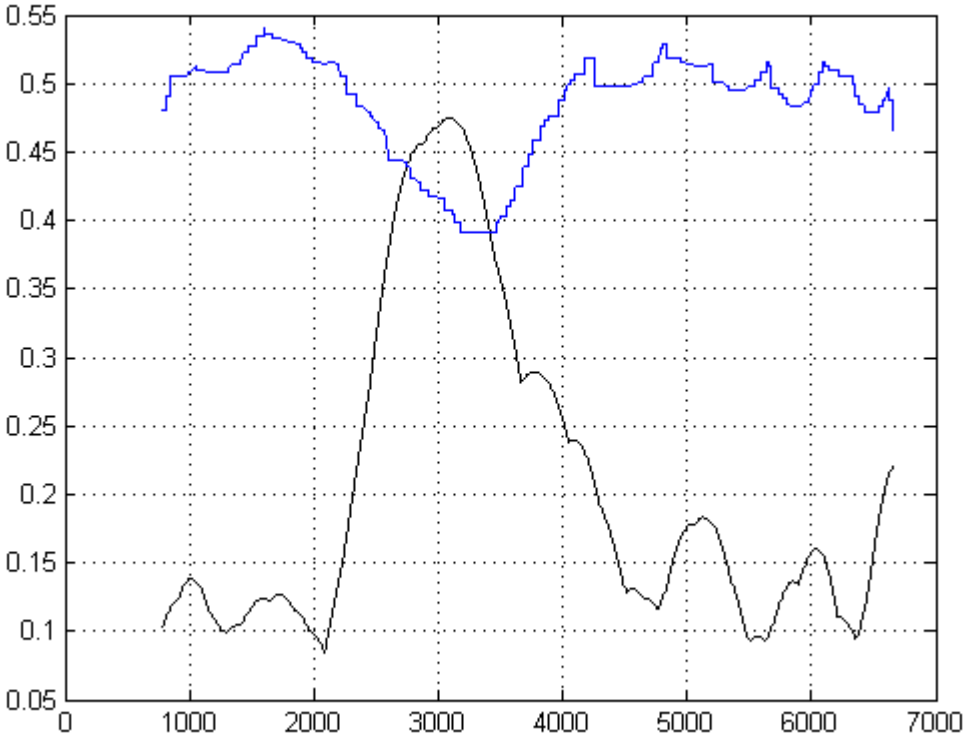
$$\Delta = \{\tilde{x} = x k_{\max} : x \in (0;1),\ \chi(x) > L(x)\}$$

— непустой *и содержит точку* $x^* k_{\max}$.

Пример синтетического набора данных



Результирующая кривая и линия уровня



САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Благодарю за внимание!

Вопросы?

«III ТМШ», Усадьба Гончаровых , 15 июня 2011 г.