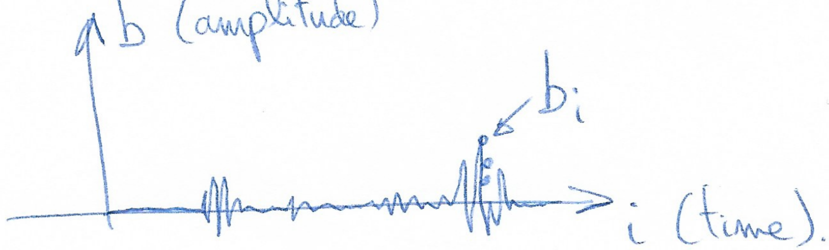
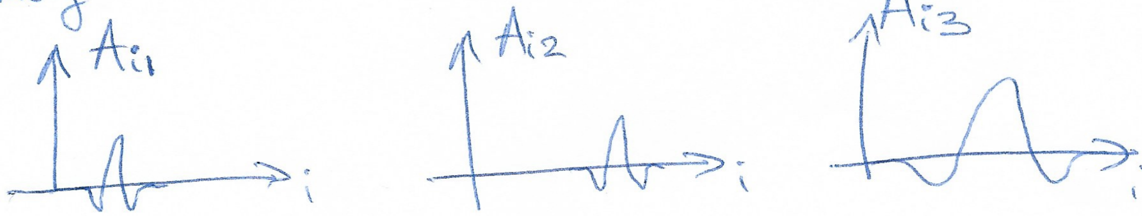
18.095 Lecture 1, Compressed sensing, 01/08/24.
Instructor: Laurent Demanet

① An example of an interesting linear system:

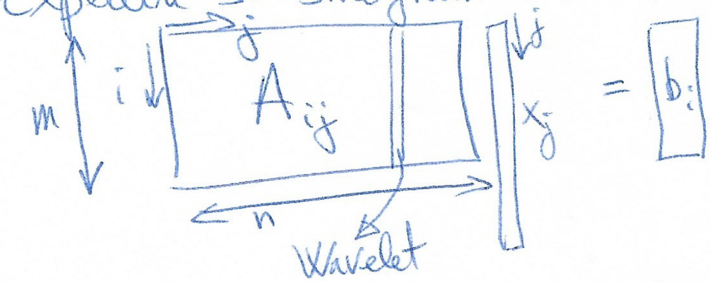
Seismogram



Dictionary of waveforms (wavelets).



Explain seismogram in terms of wavelets:



$Ax = b$.

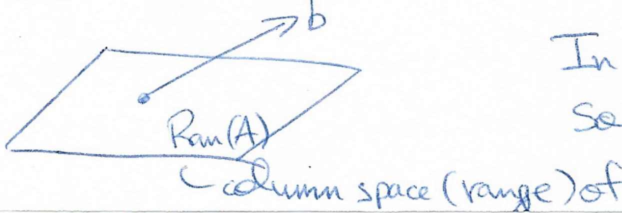
$x =$ "covariates"

But $m \le n$: under<u>determined</u>

②. Least-squares for $Ax=b$.

- Case 1 : $m=n$ $\quad$ $x = A^{-1}b$ when A is invertible. unique solution.

- Case 2 : $m \ge n$ (overdetermined).

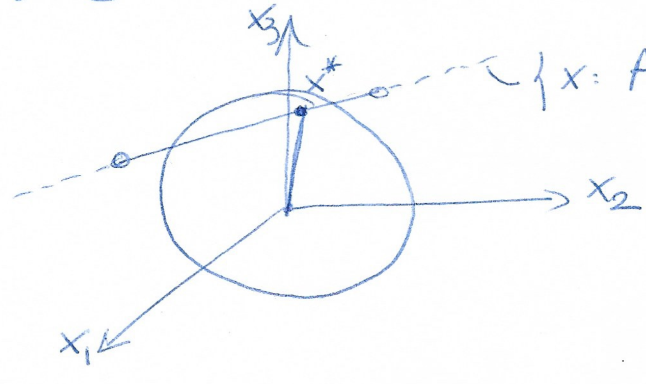


In general, $b \notin \text{Ran } A$ so there is no solution.

Least squares: $\min_x \|Ax-b\|_2^2$ (with $\|x\|_2^2 = \sum_i x_i^2$)

Solution: $x = (A^TA)^{-1}A^Tb$ (if A^TA is invertible)

- Case 3: $m \leq n$ (underdetermined).



Infinite number of solutions...

Restrict: $\min \|x\|_2 : Ax=b$.

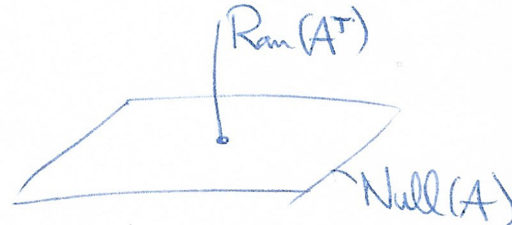
Call x^* the solution (discuss uniqueness at the end).

$x^* \perp \{x : Ax=b\}$

$\iff x^* \perp \{x : Ax=0\} = \mathrm{Null}(A)$.

i.e. $\langle x^*, x\rangle = 0 \quad \forall x \in \mathrm{Null}(A)$ (with $\langle a,b\rangle = a^Tb$).

$\iff x^* \in \mathrm{Ran}(A^T)$



i.e. $\exists y : x^* = A^Ty$

Plug in $Ax=b$: $AA^Ty = b$

$y = (AA^T)^{-1}b$ (if AA^T is invertible)

$\Rightarrow x^* = A^T(AA^T)^{-1}b$.

Rmk: x^* is not in general sparse.
i.e., it has most components non-zero.

3. Hope (unrealistic!): find <u>sparsest</u> solution

$$\min_x |\{i : x_i \neq 0\}| \quad \text{s.t.} \quad Ax = b$$

subject to

→ # of nonzero components

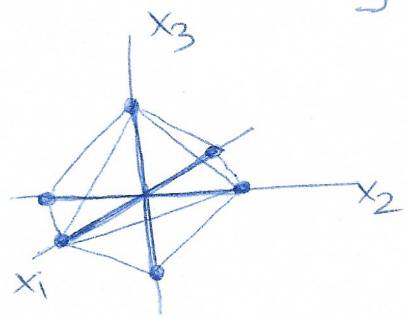
→ called $\|x\|_0$, the "ℓ_0 quasi-norm" in contrast to the ℓ_2 norm we saw earlier

$K = \|x\|_0$ is called the sparsity level of x

"x is k-sparse".

→ Combinatorially hard!

4. Next best thing: restrict to $|x_i| \leq 1$



Cross $= \{x : \|x\|_0 = 1\}$.

Convex hull of the cross: octahedron.

→ $\{x : \|x\|_1 \leq 1\}$ where $\|x\|_1 = \sum |x_i|$

(in general, $\|x\|_p = \left(\sum |x_i|^p\right)^{1/p}$ for $0<p<\infty$)

<u>hw</u>: $\|x\|_0 = \lim_{p \to 0} \|x\|_p^p$

Replace $\|x\|_0$ by its convex envelope $\|x\|_1$ (over the set $\{x : |x_i| \leq 1\}$).

and <u>hope</u> that we still get sparse solutions:

$$\boxed{\min_x \|x\|_1 : Ax = b}$$

Called LASSO / Basis pursuit. (Donoho, 1995)

⑤ How to solve?

Let $f(x) = \lambda \|x\|_1 + \|Ax - b\|_2^2$

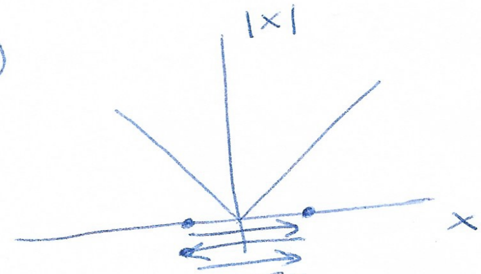
(recover LASSO in the limit $\lambda \to 0$).

$= f_1(x) \quad + f_2(x)$.

$\nabla f_2(x) = 2A^T(Ax - b)$.

$\nabla f_1(x) = \lambda \operatorname{sgn}(x)$ "sign" ± 1

ex ($n=1$)



$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0. \end{cases}$$

step size > 0

Gradient descent $x^{(k+1)} = x^{(k)} - \alpha \operatorname{sgn}(x^{(k)})$.

oscillates forever ...

Replace with
$$x^{(k+1)} = \begin{cases} x^{(k)} - \alpha & \text{if } x^{(k)} > \alpha \\ x^{(k)} + \alpha & \text{if } x^{(k)} < -\alpha \\ 0 & \text{if } |x^{(k)}| \leq \alpha \end{cases}$$

$= \text{Shrink}_\alpha(x^{(k)})$

Proximal gradient / Forward-Backward splitting

$$x^{(k+1)} = \text{Shrink}_{\lambda\alpha}\left(x^{(k)} - \alpha \nabla f_2(x^{(k)})\right).$$

Simple but slow

(Converges for α small enough when ∇f_2 is Lipschitz).

Better algorithm: look up ADMM.

Yes, this typically gives rise to sparse solutions!

(6) Compressed sensing:

$Ax = b$ with $A \in \mathbb{R}^{m \times n}$, $m \ll n$

and $A_{ij} \sim N(0,1)$ iid. (random gaussian).

$\rightarrow$ b_i = inner product with noise...

Universal compression, blind to the structure of x.

ex. $x = \Phi c$ with c k-sparse.

- x: image
- Φ: basis (wavelets, or local cosines).
- c: coefficient.

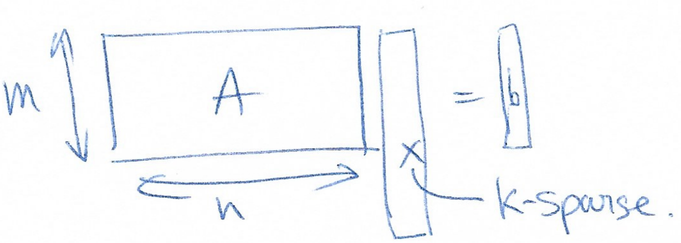
natural images: $n = 10^6$ to 10^7.
$k = 2.5 \cdot 10^4$ wavelet coeff.

ℓ_1 "decoder":

$$\min_c \|c\|_1 \text{ s.t. } A\Phi c = b. \qquad (P_1)$$

Assume (wlog, for simplicity) Φ = identity

Compressed sensing question: if $b = Ax_0$ for some k-sparse x_0, can (P_1) recover $x = x_0$ as minimizer?



Need $m \geq k$

Hope $m \ll n$.

<u>Theorem</u> (2004, Candès-Romberg-Tao / Donoho)

Let x_0 be k-sparse : $\|x\|_0 = k$.
Let $b = Ax_0$ for $A \in \mathbb{R}^{m\times n}$, $m \le n$,
$A_{ij} \sim N(0,1)$ iid.

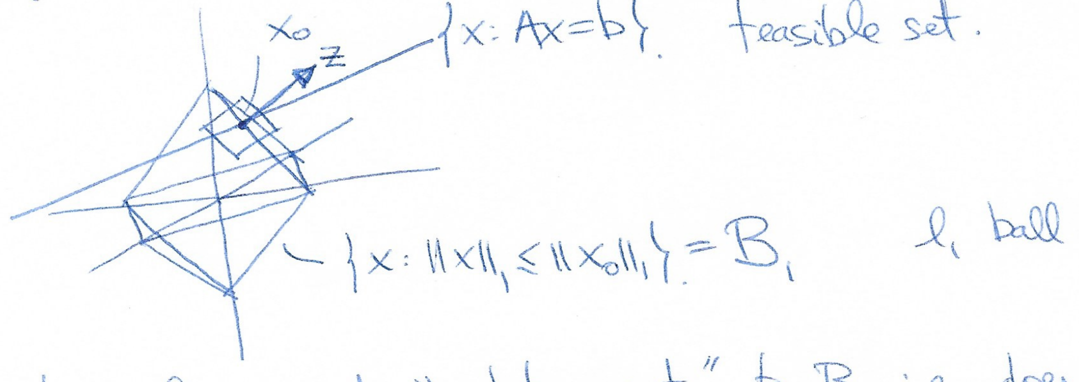
then (P_1) recovers x_0 with high probability provided

$$m \geq C k \log(n) \quad \text{for some } C>0.$$

Furthermore, the solution of (P_1) coïncides with that of ℓ_0 minimization.

In practice, $m \simeq 5k$ suffices. (Whp means $1-ce^{-cm}$).

Proof idea:



Show feasible set "subtangent" to B, i.e., does not cross it.

(codim-1) Plane (x_0, z) such that:
- contains feasible set
- entirely on one side of B.

base point / normal

→ construct z and show these 2 conditions are satisfied.

z called a <u>dual certificate</u>.