# Intro to wavelets part 1

Recall, we considered the sparsity promoting functional:

$$\bar{x} = \underset{x}{\arg\min} \left\{ \|Ax-b\|_2^2 + 2\tau \|x\|_1 \right\}$$

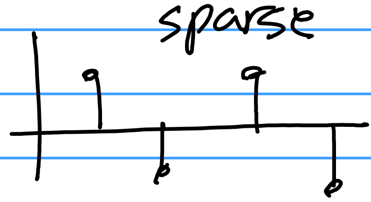
Can be minimized by ISTA:

$$x^{n+1} = S_\tau \left[ x^n + A^T b - A^T A x^n \right]$$

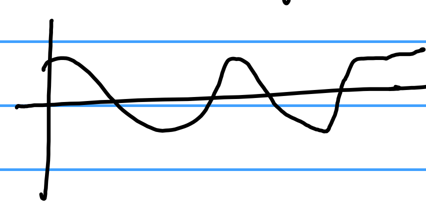
for any $x^0$ as long as $\|A\|_1$.

Typically, solution $x$ to $Ax \approx b$ is not necessarily expected to be sparse.

sparse



non-sparse



Would be nice to have $W$ and $W^{-1}$ s.t.

$WW^{-1} = I$, $\quad w = W \cdot x$ is sparse

Then it makes sense to consider

$$\bar{w} = \underset{w}{\arg\min} \left\{ \|Ax-b\|^2 + 2\tau \|w\|_1 \right\}$$

$$= \underset{w}{\arg\min} \left\{ \|AW^{-1}w - b\|^2 + 2\tau \|w\|_1 \right\}$$

# Haar Wavelet and Transform

Discrete signal: $f = (f_1, f_2, \ldots, f_N)$

let $N$ be even and $> 0$.

$f_1 = g(t_1),\ f_2 = g(t_2) \ldots,\ f_N = g(t_N)$

The Haar transform decomposes a discrete signal into two signals of half its length.

$a^1 = (a_1, a_2, \ldots, a_{N/2})$ trend subsignal

$$a_m = \frac{f_{2m-1} + f_{2m}}{\sqrt{2}}$$

$d^1 = (d_1, d_2, \ldots, d_{N/2})$ first fluctuation

take running difference between two values and multiply by $\sqrt{2}$:

$$d_m = \frac{f_{2m-1} - f_{2m}}{\sqrt{2}}$$

Ex) $f = (4, 6, 10, 12, 8, 6, 5, 5)$

$$\begin{array}{cccccccc}
4 & 6 & 10 & 12 & 8 & 6 & 5 & 5 \\
\multicolumn{2}{c}{\searrow\ \swarrow} & \multicolumn{2}{c}{\searrow\ \swarrow} & \multicolumn{2}{c}{\searrow\ \swarrow} & \multicolumn{2}{c}{\searrow\ \swarrow} \\
\multicolumn{2}{c}{-1} & \multicolumn{2}{c}{-1} & \multicolumn{2}{c}{1} & \multicolumn{2}{c}{0} \\
\multicolumn{2}{c}{\downarrow} & \multicolumn{2}{c}{\downarrow} & \multicolumn{2}{c}{\downarrow} & \multicolumn{2}{c}{\downarrow} \\
\multicolumn{2}{c}{-\sqrt{2}} & \multicolumn{2}{c}{-\sqrt{2}} & \multicolumn{2}{c}{\sqrt{2}} & \multicolumn{2}{c}{0}
\end{array}$$

$d^{(1)}$: $-\sqrt{2} \quad -\sqrt{2} \quad \sqrt{2} \quad 0$

detail coefficients

$$\begin{array}{cccccccc}
4 & 6 & 10 & 12 & 8 & 6 & 5 & 5 \\
\multicolumn{2}{c}{\searrow\ \swarrow} & \multicolumn{2}{c}{\searrow\ \swarrow} & \multicolumn{2}{c}{\searrow\ \swarrow} & \multicolumn{2}{c}{\searrow\ \swarrow} \\
\multicolumn{2}{c}{5} & \multicolumn{2}{c}{11} & \multicolumn{2}{c}{7} & \multicolumn{2}{c}{5} \\
\multicolumn{2}{c}{\downarrow} & \multicolumn{2}{c}{\downarrow} & \multicolumn{2}{c}{\downarrow} & \multicolumn{2}{c}{\downarrow} \\
\multicolumn{2}{c}{5\sqrt{2}} & \multicolumn{2}{c}{11\sqrt{2}} & \multicolumn{2}{c}{7\sqrt{2}} & \multicolumn{2}{c}{5\sqrt{2}}
\end{array}$$

$a^{(1)}$: $5\sqrt{2} \quad 11\sqrt{2} \quad 7\sqrt{2} \quad 5\sqrt{2}$

$$f \xmapsto{H_1} (a^1 \mid d^1)$$

$$(4, 6, 10, 12, 8, 6, 5, 5) \xmapsto{H_1} (5\sqrt{2}, 11\sqrt{2}, 7\sqrt{2}, 5\sqrt{2} \mid -\sqrt{2}, -\sqrt{2}, \sqrt{2}, 0)$$

The nice thing is that we can easily map the transform back to $f$:

$$f = \left( \frac{a_1 + d_1}{\sqrt{2}}, \frac{a_1 - d_1}{\sqrt{2}}, \dots, \frac{a_{N/2} + d_{N/2}}{\sqrt{2}}, \frac{a_{N/2} - d_{N/2}}{\sqrt{2}} \right)$$

$a^{(1)}$: $5\sqrt{2}, \quad 11\sqrt{2}, \quad 7\sqrt{2}, \quad 5\sqrt{2}$.

$d^{(1)}$: $-\sqrt{2}, \quad -\sqrt{2}, \quad \sqrt{2}, \quad 0$

4 6    10 12    8 6    5 5

The magnitudes of $d'$ often smaller than of $f$.

Notice Energy Conservation:

$$a_1^2 + d_1^2 = \left[\frac{f_1 + f_2}{\sqrt{2}}\right]^2 + \left[\frac{f_1 - f_2}{\sqrt{2}}\right]^2 =$$

$$= \frac{f_1^2 + 2f_1f_2 + f_2^2}{2} + \frac{f_1^2 - 2f_1f_2 + f_2^2}{2} = f_1^2 + f_2^2$$

Generally: $\|a^{(1)}\|^2 + \|d^{(1)}\|^2 = \|f\|^2$

If we define $\tilde{W}_1^2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0, \dots, 0\right)$

$$\Rightarrow d^{(1)} = \frac{f_1 - f_2}{\sqrt{2}} = f \cdot \tilde{W}_1^2$$

Notice that $\tilde{w}_2 = \left(0, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right)$ translation of $\tilde{w}_1$

Also note that $\tilde{S}_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \dots, 0\right)$

$\tilde{S}_2 = \left(0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \dots, 0\right)$ $\Rightarrow a_1 = \tilde{S}_1 \cdot f$ and so on

$a_2 = \tilde{S}_2 f$

We can simplify all these by dropping the $\sqrt{2}$ factor:

Define Haar scaling function as: $\phi(x) = \begin{cases} 1, & 0 \le x < 1 \\ 0, & \text{otherwise} \end{cases}$

Define Haar Wavelet function as:

$$\psi(x) = \phi(2x) - \phi(2x-1) = \begin{cases} 1, & 0 \le x < \frac{1}{2} \\ -1, & \frac{1}{2} \le x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Then the set of functions $\{2^{j/2}\phi(2^j x - k) ; k \in \mathbb{Z}\}$ is an orthonormal basis.

$\underline{j=0}$ : $\int_{-\infty}^{+\infty} \phi^2(x)\, dx = \int_0^1 1 \cdot dx = 1$

$\underline{j=1}$ : $2^{\frac{1}{2}} \phi(2x)$ $\qquad \phi(2x) = \begin{cases} 1, & 0 < x < \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$

$= \sqrt{2}\, \phi(2x)$

$$\Rightarrow \int_{-\infty}^{+\infty} \left[\sqrt{2}\,\phi(2x)\right]^2 dx = \int_0^{1/2} (\sqrt{2} \cdot 1)^2 dx = 1$$

$$f = \left(\frac{a_1 + d_1}{\sqrt{2}}, \frac{a_1 - d_1}{\sqrt{2}}, \dots, \frac{a_{N/2} - d_{N/2}}{\sqrt{2}}, \frac{a_{N/2} - d_{N/2}}{\sqrt{2}}\right)$$

$$= \left(\frac{a_1}{\sqrt{2}}, \frac{a_1}{\sqrt{2}}, \dots, \frac{a_{N/2}}{\sqrt{2}}, \frac{a_{N/2}}{\sqrt{2}}\right) + \left(\frac{d_1}{\sqrt{2}}, \frac{-d_1}{\sqrt{2}}, \dots, \frac{d_{N/2}}{\sqrt{2}}, \frac{-d_{N/2}}{\sqrt{2}}\right)$$

$$= \left(a_1 \tilde{S}_1^1 + \dots + a_{N/2} \tilde{S}_{N/2}^1\right) + \left(d_1 \tilde{W}_1^1 + \dots + d_{N/2} \tilde{W}_{N/2}^1\right)$$

$$= (f\cdot\tilde{S}_1')\tilde{S}_1' + \dots + (f\cdot\tilde{S}_{N/2}')\tilde{S}_{N/2}' + (f\cdot\tilde{W}_1')\tilde{W}_1' + \dots + (f\cdot\tilde{W}_{N/2}')\tilde{W}_{N/2}'$$

where $\tilde{S}_1' = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \dots, 0\right)$ ; $\tilde{W}_1' = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, \dots, 0\right)$

$\tilde{S}_2' = \left(0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \dots, 0\right)$ ; $\tilde{W}_2 = \left(0, 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, \dots, 0\right)$

$$f \xrightarrow{\text{Haar Transform}} \left(a^{(1)} \,|\, d^{(1)}\right)$$

But we can perform the transform again on $a^{(1)}$:

$$f \xrightarrow{\text{two level Haar}} \left(\underbrace{a^{(2)}}_{\substack{\text{small amount}\\ \text{of scaling}\\ \text{"approx." coefficients}}} \,|\, \underbrace{d^{(2)} \,|\, d^{(1)}}_{\substack{\text{large amount of}\\ \text{smaller (in magnitude)}\\ \text{detail coefficients}}}\right)$$

Now let's go back to the optimization problem:

$$\left(\bar{w} = \underset{w}{\operatorname{argmin}} \left\{ \|AW^{-1}w - b\|^2 + 2\tau\|w\|_1 \right\}\right) \quad \|w\|_1 + \|x\|_1$$

set $M = AW^{-1}$. Then can use e.g. FISTA:

with $T(w) = S_\tau\left[w + M^T\cdot b - M^T M w\right]$

and $w^{n+1} = T\left(w^n + \frac{t_n - 1}{t_{n+1}}\left(w^n - w^{n-1}\right)\right)$

$t_1 = 1$ and $t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2}$ (See Beck and Teboulle FISTA paper)

But we must have $\|M\| \leq 1$. Recall majorization-min. derivation.

spectral norm

To find $\|M\|$ = largest $|\sigma_i|$ of $M$, we can use power iteration on the matrix $M^T M$.

Let $B = M^T M$ then $B$ has eigenvalues

$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$

If we multiply a vector $x_0 \in \mathbb{R}^n$ with $B$:

$$x^{(k)} = B x^{(k-1)} = B^k x^{(0)}, \quad k = 1, 2, \dots$$

Any vector in $\mathbb{R}^n$ can be written in terms of the eigenvectors of $B$:

$$x^{(0)} = \alpha_1 x_1 + \dots + \alpha_n x_n$$

$$x^{(k)} = B^k x^{(0)} = \sum_{i=1}^{n} \alpha_i B^k x_i$$

$$= \sum_{i=1}^{n} \alpha_i \lambda_i^k x_i$$

$B x_i = \lambda_i x_i$

$B^k x_i = \lambda_i^k x_i$

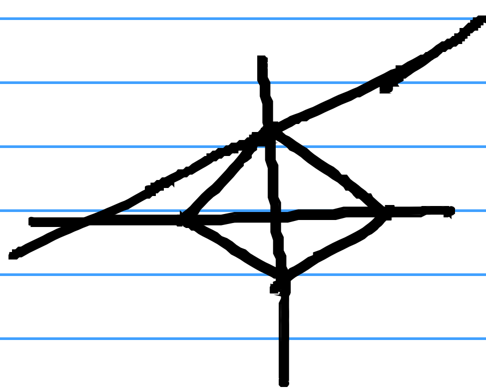
$$= \left[\alpha_1 \lambda_1^k x_1 + \alpha_2 \lambda_2^k x_2 + \dots + \alpha_n \lambda_n^k x_n\right]$$

$$= \lambda_1^k \left[\alpha_1 x_1 + \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i\right]$$

$\left|\frac{\lambda_i}{\lambda_1}\right| < 1$

$\rightarrow \lambda_1^k \alpha_1 x_1$ (direction of $x^k$ converges to

$= \mathcal{Z} x_1$ that of $x_1$, the dominant eigenvector).



$y = mx - b$

$\|x\|_1$

Review of sparse regularization $Ax \approx b$ and $x$ is expected to be sparse

$$F_\tau(x) = \|Ax-b\|^2 + 2\tau\|x\|_1$$

relevant when vector $x$ is sparse.

We are using a wavelet basis representation:

$Wx = w$. Notice that for orthogonal transform $W^{-1} = W^T$ so that:

$$\|Wx\|^2 = \|w\|^2 = \langle Wx, Wx\rangle = \langle x, W^T W x\rangle$$
$$= \langle x, Ix\rangle = \langle x, x\rangle = \|x\|^2$$

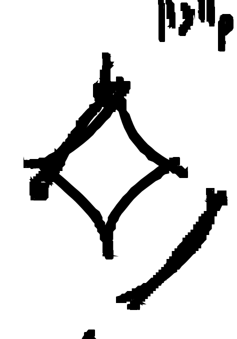
(i.e. "energy" is preserved)

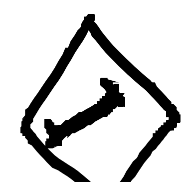
Hence, if one considers e.g.

$\min\{\|Ax-b\|^2 + \lambda\|w\|^2\}$ then the result $\boxed{\ell^1}$ $\|x\|_p^p$

is the same as the original problem:

$\min\{\|Ax-b\|^2 + \lambda\|x\|^2\}$





Thus, sparsity promoting penalties are commonly used with wavelet basis.

For any $\alpha \in (0,1)$:

$$\|\alpha x + (1-\alpha) x\|_1 \leq \|\alpha x\|_1 + \|(1-\alpha) x\|_1$$
$$= \alpha \|x\|_1 + (1-\alpha)\|x\|_1$$

Hence $F_\tau(x)$ is convex. A local min is a global min.

## Optimality conditions

$$\begin{cases} [A^T(b - Ax)]_k = \tau \, \mathrm{sgn}(x_k) \ ; \ \forall k \ x_k \neq 0 \\ |[A^T(b - Ax)]_k| \leq \tau \ ; \ , \ \forall k \ x_k = 0 \end{cases}$$

$\vec{x} = 0$ is a solution for $\tau \geq \underline{\underline{\|A^T b\|_\infty}}$.

Local min is global min: (for any convex function)

suppose $f(x) \leq f(z)$ for small $\varepsilon$ $|x - z| \leq \varepsilon$

$$f(x) \leq \boxed{f(ty + (1-t)x)} = f(x+\varepsilon) \text{ for small } t$$
$$\leq t f(y) + (1-t) f(x) = tf(y) + f(x) - tf(x)$$

$\Rightarrow$ $\Rightarrow$ $tf(x) \leq tf(y) \Rightarrow \boxed{f(x) \leq f(y)}$

Since $\nabla F_\tau(x)$ does not exist, need to use more clever methods

$$G(x,y) \geq F(x) \quad ; \quad G(x,x) = F(x)$$

$$x^{(n+1)} = \underset{x}{\arg\min}\ G(x, x^n)$$

$$F(x^{n+1}) = G(x^{n+1}, x^{n+1}) \leq G(x^{n+1}, x^n)$$
$$\leq G(x^n, x^n) = F(x^n, x^n)$$

$$G(x,y) = \|Ax-b\|^2 - \|A(x-y)\|^2 + \|x-y\|^2 + 2\tau\|x\|_1$$

need $G(x,y) \geq F(x)$
$$= \|Ax-b\|^2 + 2\tau\|x\|_1$$

$$\Rightarrow \|x-y\|^2 - \|A(x-y)\|^2 \geq 0$$
$$\Rightarrow \|A(x-y)\|^2 \leq \|x-y\|^2$$

$$\Rightarrow \|A(x-y)\|^2 \le \|A\|^2 \|x-y\|^2$$

$$\hookrightarrow \|A\|^2 \le 1 \Rightarrow \boxed{\|A\| \le 1}$$

$$x^{n+1} = \underset{x}{\operatorname{argmin}}\ G(x, x^n) = \dots =$$

$$= S_\tau\left[x^n + A^T b - A^T A x^n\right]$$

ISTA slow converging

$$\left| F(x^n) - F(\bar{x}) \right| \sim O\left(\frac{1}{n}\right)$$

global minimizer

FISTA (speed up trick by Nesterov 1983)

$$M = AW^{-1} \quad \left\{ \|Mw - b\|^2 + 2\tau \|w\|_1 \right\}$$

need $M$ to have $\|M\| < 1$

$$\tilde{M} = \frac{1}{|\sigma_1|} M \ ; \ \tilde{\tau} = \frac{b}{|\sigma_1|}$$

where $\sigma_1 = \max \text{ s.v.}(M)$

$$\Rightarrow \left\{ \| \tilde{M} w - \tilde{b} \|^2 + 2\tau \|w\|_1 \right\}$$

$$w^{n+1} = S_\tau \left[ w^n + \tilde{M}^T \tilde{b} - \tilde{M}^T \tilde{M} w^n \right]$$

$$= S_\tau \left[ w^n + \frac{1}{|\sigma_1|^2} M^T b - \frac{1}{|\sigma_1|^2} M^T M w^n \right]$$

forming $\underset{(m\times n)}{M} = A \underset{(n\times n)}{W^{-1}}$ is expensive!

estimate $\sigma_1$ = max s.v. of $M = AW^{-1}$

$$M^T M = (U \Sigma V^T)^T (U \Sigma V^T)$$

$$= V \Sigma \underbrace{U^T U}_{=I} \Sigma V^T = \underbrace{V \Sigma^2 V^T}_{\text{eigendecomp.}}$$

the largest s.v. of $M$ is the square root of largest e-val of $M^T M$

$$M^T M v = (AW^{-1})^T (AW^{-1}) v$$

$$= \underline{W^{-T}} A^T A W^{-1} v \qquad W^T = W^{-1} \rightarrow W^{-T} = W!$$

$M^T M v = W A^T A W^{-1} v$

routine to form $W^{-1} v$

$y = W^{-1} v \quad Ay \Rightarrow A^T(Ay) = z$

apply forward $W^T$ routine on $z$

$M = \underbrace{A}_{m \times n} \underbrace{W^{-1}}_{n \times n}$. Need to compute

$Mx = AW^{-1}x = A(W^{-1}x)$

notice if $W$ non-orthogonal then $W^{-T} \neq W$.

$M^T y = (AW^{-1})^T y = W^{-T} A^T y = W A^T y$

since $W^{-1} = W^T$ for orthogonal transforms

## QR factorization

$A = QR$ (unpivoted QR)

$AP = QR$ (column pivoted QR)

$P$ is a permutation matrix of the columns of $A \Rightarrow A = QRP^T$ ($P$ is orthogonal matrix)