Weighted Least Squares

Consider weighted $\ell_2$ minimization.
Weighted Least Squares

- Consider weighted $\ell_2$ minimization.
- Let $w_j > 0, \ j = 1, \ldots, N$ be a positive weight
Weighted Least Squares

- Consider weighted $\ell_2$ minimization.
- Let $w_j > 0, j = 1, \ldots, N$ be a positive weight
- $\|u\|_{\ell_2(w)} := \left[ \sum_{j=1}^{N} w_j u_j^2 \right]^{1/2}$
Weighted Least Squares

- Consider weighted $\ell_2$ minimization.
- Let $w_j > 0$, $j = 1, \ldots, N$ be a positive weight
- $\|u\|_{\ell_2(w)} := \left[ \sum_{j=1}^{N} w_j u_j^2 \right]^{1/2}$
- $\langle u, v \rangle_w := \sum_{j=1}^{N} w_j u_j v_j$
Weighted Least Squares

- Consider weighted $\ell_2$ minimization.
- Let $w_j > 0$, $j = 1, \ldots, N$ be a positive weight.

\[ \|u\|_{\ell_2(w)} := \left[ \sum_{j=1}^{N} w_j u_j^2 \right]^{1/2} \]

\[ \langle u, v \rangle_w := \sum_{j=1}^{N} w_j u_j v_j \]

Define $x(w) := \text{Argmin}_{z \in \mathcal{F}(y)} \|z\|_{\ell_2(w)}$
Weighted Least Squares

- Consider weighted $\ell_2$ minimization.

- Let $w_j > 0$, $j = 1, \ldots, N$ be a positive weight

- $\|u\|_{\ell_2(w)} := \left[ \sum_{j=1}^{N} w_j u_j^2 \right]^{1/2}$

- $\langle u, v \rangle_w := \sum_{j=1}^{N} w_j u_j v_j$

- Define $x(w) := \text{Argmin}_{z \in \mathcal{F}(y)} \|z\|_{\ell_2(w)}$

- $x(w) = x - \eta(w)$ where $\eta(w) := \text{Argmin}_{\eta \in \mathcal{N}} \|x - \eta\|_{\ell_2(w)}$
Weighted Least Squares

Consider weighted $\ell_2$ minimization.

Let $w_j > 0$, $j = 1, \ldots, N$ be a **positive** weight.

$$\|u\|_{\ell_2(w)} := \left[ \sum_{j=1}^{N} w_j u_j^2 \right]^{1/2}$$

$$\langle u, v \rangle_w := \sum_{j=1}^{N} w_j u_j v_j$$

Define $x(w) := \text{Argmin}_{z \in \mathcal{F}(y)} \|z\|_{\ell_2(w)}$

$x(w) = x - \eta(w)$ where $\eta(w) := \text{Argmin}_{\eta \in \mathcal{N}} \|x - \eta\|_{\ell_2(w)}$

Note that this solution is characterized by the orthogonality conditions $\langle x(w), \eta \rangle_w = 0$, $\eta \in \mathcal{N}$
The Algorithm

\[ J(z, w, \epsilon) := \frac{1}{2} \left[ \sum_{j=1}^{N} z_j^2 w_j + \sum_{j=1}^{N} (\epsilon^2 w_j + w_j^{-1}) \right] . \]
The Algorithm

\[ J(z, w, \epsilon) := \frac{1}{2} \left[ \sum_{j=1}^{N} z_j^2 w_j + \sum_{j=1}^{N} (\epsilon^2 w_j + w_j^{-1}) \right]. \]

Initialize: \( w^0 := (1, \ldots, 1), \quad \epsilon_0 := 1 \)
The Algorithm

\[ \mathcal{J}(z, w, \epsilon) := \frac{1}{2} \left[ \sum_{j=1}^{N} z_j^2 w_j + \sum_{j=1}^{N} (\epsilon^2 w_j + w_j^{-1}) \right]. \]

Initialize: \( w^0 := (1, \ldots, 1), \quad \epsilon_0 := 1 \)

\( x^{m+1} := \text{Argmin}_{z \in \mathcal{F}(y)} \mathcal{J}(z, w^m, \epsilon_m), \quad m = 0, 1, \ldots \)
The Algorithm

\[ J(z, w, \epsilon) := \frac{1}{2} \left[ \sum_{j=1}^{N} z_j^2 w_j + \sum_{j=1}^{N} (\epsilon^2 w_j + w_j^{-1}) \right]. \]

Initialize:  
\[ w^0 := (1, \ldots, 1), \quad \epsilon_0 := 1 \]

\[ x^{m+1} := \text{Argmin}_{z \in \mathcal{F}(y)} J(z, w^m, \epsilon_m), \quad m = 0, 1, \ldots \]

\[ \epsilon_{m+1} := \min(\epsilon_m, \frac{r(x^{m+1})_K}{N}), \]
The Algorithm

\[ J(z, w, \epsilon) := \frac{1}{2} \left[ \sum_{j=1}^{N} z_j^2 w_j + \sum_{j=1}^{N} (\epsilon^2 w_j + w_j^{-1}) \right] . \]

Initialize: \( w^0 := (1, \ldots, 1), \quad \epsilon_0 := 1 \)

\( x^{m+1} := \text{Argmin}_{z \in \mathcal{F}(y)} J(z, w^m, \epsilon^m), \quad m = 0, 1, \ldots \)

\( \epsilon_{m+1} := \min(\epsilon_m, \frac{r(x^{m+1})_K}{N}) \),

\( w^{m+1} := \text{Argmin}_{w > 0} J(x^{m+1}, w, \epsilon_{m+1}) \)
The Algorithm

\[ J(z, w, \epsilon) := \frac{1}{2} \left[ \sum_{j=1}^{N} z_j^2 w_j + \sum_{j=1}^{N} (\epsilon^2 w_j + w_j^{-1}) \right]. \]

Initialize: \( w^0 := (1, \ldots, 1), \quad \epsilon_0 := 1 \)

\( x^{m+1} := \text{Argmin}_{z \in F(y)} J(z, w^m, \epsilon_m), \quad m = 0, 1, \ldots \)

\( \epsilon_{m+1} := \min(\epsilon_m, \frac{r(x^{m+1})_K}{N}) \),

\( w^{m+1} := \text{Argmin}_{w > 0} J(x^{m+1}, w, \epsilon_{m+1}) \)

\( w_j^{m+1} = [ (x_j^{m+1})^2 + \epsilon_{m+1}^2 ]^{-1/2} \)
The Algorithm

\[ J(z, w, \epsilon) := \frac{1}{2} \left[ \sum_{j=1}^{N} z_j^2 w_j + \sum_{j=1}^{N} (\epsilon^2 w_j + w_j^{-1}) \right] . \]

Initialize: \( w^0 := (1, \ldots, 1) \), \( \epsilon_0 := 1 \)

\( x^{m+1} := \text{Argmin}_{z \in \mathcal{F}(y)} J(z, w^m, \epsilon_m), \quad m = 0, 1, \ldots \)

\( \epsilon_{m+1} := \min(\epsilon_m, \frac{r(x^{m+1})_K}{N}) \),

\( w^{m+1} := \text{Argmin}_{w > 0} J(x^{m+1}, w, \epsilon_{m+1}) \)

\( w_j^{m+1} = \left[ (x_j^{m+1})^2 + \epsilon_{m+1}^2 \right]^{-1/2} \)

If \( \epsilon_{m+1} = 0 \) stop algorithm: \( x^{m+1} \) is \( K \)-sparse
Convergence Theorem

Theorem
Let \( k \geq 1 \) and define \( K = k + 6 \). We assume that \( \Phi \) satisfies the Null Space Property for \( \ell_1 \) of order \( 3K \) for \( \gamma \leq 1/2 \). Let \( x^* \) be the unique minimum \( \ell_1 \) minimizer from \( F(y) \). Then, for each \( y \in \mathbb{R}^n \), the Algorithm converges and its limit \( \bar{x} \) satisfies

\[
\| x^* - \bar{x} \|_{\ell_1} \leq C_1 \sigma_k(x^*)_{\ell_1}, \quad C_1 := \frac{5(1 + \gamma)}{1 - \gamma}.
\]

In particular if \( x^* \) is \( k \)-sparse then \( x^m \) converges to \( x^* \).
We define \( r(z) \) as the rearrangement of the sequence \( |z_j| \) into decreasing order. In other words \( r(z)_j \) is the \( j \)-th largest of the \( |z_\nu| \)
We define $r(z)$ as the rearrangement of the sequence $|z_j|$ into decreasing order. In other words, $r(z)_j$ is the $j$-th largest of the $|z_\nu|$. Rearrangement is a Lipschitz map on $\| \cdot \|_{L_\infty}$. 
We define \( r(z) \) as the rearrangement of the sequence \( |z_j| \) into decreasing order. In other words \( r(z)_j \) is the \( j \)-th largest of the \( |z_\nu| \).

Rearrangement is a Lipschitz map on \( \| \cdot \|_{\ell_\infty} \).

More precisely \( \| r(z) - r(z') \|_{\ell_\infty} \leq \| z - z' \|_{\ell_\infty} \).
We define $r(z)$ as the rearrangement of the sequence $|z_j|$ into decreasing order. In other words $r(z)_j$ is the $j$-th largest of the $|z_\nu|$.

Rearrangement is a Lipschitz map on $\| \cdot \|_{\ell_\infty}$.

More precisely $\| r(z) - r(z') \|_{\ell_\infty} \leq \| z - z' \|_{\ell_\infty}$.

Moreover, for any $j$, we have

$$\left| \sigma_j(z)_{\ell_1} - \sigma_j(z')_{\ell_1} \right| \leq \| z - z' \|_{\ell_1}$$
We define $r(z)$ as the rearrangement of the sequence $|z_j|$ into decreasing order. In other words $r(z)_j$ is the $j$-th largest of the $|z_\nu|$

Rearrangement is a Lipschitz map on $\| \cdot \|_{\ell_\infty}$

More precisely $\|r(z) - r(z')\|_{\ell_\infty} \leq \|z - z'\|_{\ell_\infty}$

Moreover, for any $j$, we have

$$|\sigma_j(z)_{\ell_1} - \sigma_j(z')_{\ell_1}| \leq \|z - z'\|_{\ell_1}$$

For any $J > j$, we have

$$(J - j)r(z)_J \leq \|z - z'\|_{\ell_1} + \sigma_j(z')_{\ell_1}$$
The Geometric Property

Assume that NSP holds for some $k$ and $\gamma < 1$
The Geometric Property

- Assume that NSP holds for some $k$ and $\gamma < 1$
- For any $z, z' \in \mathcal{F}(y)$

$$
\|z' - z\|_{\ell_1} \leq \frac{1 + \gamma}{1 - \gamma} (\|z'\|_{\ell_1} - \|z\|_{\ell_1} + 2\sigma_k(z)_{\ell_1}).
$$
Convergence of Algorithm

The proof of convergence of the algorithm is not direct
Convergence of Algorithm

- The proof of convergence of the algorithm is not direct.
- We first show that $x^m$ are bounded.
Convergence of Algorithm

- The proof of convergence of the algorithm is not direct.
- We first show that $x^m$ are bounded.
- Then we show each convergent subsequence has the same limit.
Properties of the $x^m$

The starting point is the following monotonicity of $\mathcal{J}$

$$\mathcal{J}(x^{m+1}, w^{m+1}, \epsilon_{m+1}) \leq \mathcal{J}(x^{m+1}, w^{m}, \epsilon_{m+1})$$

$$\leq \mathcal{J}(x^{m+1}, w^{m}, \epsilon_{m}) \leq \mathcal{J}(x^{m}, w^{m}, \epsilon_{m})$$
Properties of the $x^m$

- The starting point is the following monotonicity of $\mathcal{J}$
  \[ \mathcal{J}(x^{m+1}, w^{m+1}, \epsilon_{m+1}) \leq \mathcal{J}(x^{m+1}, w^m, \epsilon_{m+1}) \]
  \[ \leq \mathcal{J}(x^{m+1}, w^m, \epsilon_m) \leq \mathcal{J}(x^m, w^m, \epsilon_m) \]

- From this we get that the $x^m$ are bounded
  \[ \|x^m\|_{\ell_1} \leq \mathcal{J}(x^0, w^0, \epsilon_0) =: C_0 \]
Properties of the $x^m$

- The starting point is the following monotonicity of $\mathcal{J}$
  \[
  \mathcal{J}(x^{m+1}, w^{m+1}, \epsilon_{m+1}) \leq \mathcal{J}(x^{m+1}, w^{m}, \epsilon_{m+1}) \\
  \leq \mathcal{J}(x^{m+1}, w^{m}, \epsilon_{m}) \leq \mathcal{J}(x^{m}, w^{m}, \epsilon_{m})
  \]

- From this we get that the $x^m$ are bounded
  \[
  \|x^m\|_{\ell_1} \leq \mathcal{J}(x^0, w^0, \epsilon_0) =: C_0
  \]

- Indeed \[
  \|x^m\|_{\ell_1} \leq \sum_{j=1}^{N} [(x_j^m)^2 + \epsilon_m^2]^{1/2} = \mathcal{J}(x^m, w^m, \epsilon_m).
  \]
Properties of the $x^m$

- The starting point is the following monotonicity of $\mathcal{J}$

$$\mathcal{J}(x^{m+1}, w^{m+1}, \epsilon_{m+1}) \leq \mathcal{J}(x^{m+1}, w^{m}, \epsilon_{m+1})$$

$$\leq \mathcal{J}(x^{m+1}, w^{m}, \epsilon_{m}) \leq \mathcal{J}(x^{m}, w^{m}, \epsilon_{m})$$

- From this we get that the $x^m$ are bounded

$$\|x^m\|_{\ell^1} \leq \mathcal{J}(x^{0}, w^{0}, \epsilon_{0}) =: C_0$$

- Indeed

$$\|x^m\|_{\ell^1} \leq \sum_{j=1}^{N} [(x_j^m)^2 + \epsilon_m^2]^{1/2} = \mathcal{J}(x^{m}, w^{m}, \epsilon_{m})$$

- and the weights are bounded from below: for each $m$

$$w_j^m \geq A^{-1}, \quad A := C_0 + \epsilon_0, \quad j = 1, \ldots, N.$$
Given any $y \in \mathbb{R}^n$, the $x^m$ satisfy

$$
\sum_{m=1}^{\infty} \|x^{m+1} - x^m\|_2^2 \leq 2 AJ(x^0, w^0, \epsilon_0)
$$
Key Lemma

Given any \( y \in \mathbb{R}^n \), the \( x^m \) satisfy

\[
\sum_{m=1}^{\infty} \| x^{m+1} - x^m \|_{\ell_2}^2 \leq 2AJ(x^0, w^0, \epsilon_0)
\]

Here \( A \) is the constant in the weight inequality.
Key Lemma

- Given any \( y \in \mathbb{R}^n \), the \( x^m \) satisfy

\[
\sum_{m=1}^{\infty} \|x^{m+1} - x^m\|_{\ell_2}^2 \leq 2AJ(x^0, w^0, \epsilon_0)
\]

- Here \( A \) is the constant in the weight inequality.

- For each \( m = 1, 2, \ldots \), we have

\[
2[J(x^m, w^m, \epsilon_m) - J(x^{m+1}, w^m, \epsilon_m)]
= \langle x^m, x^m \rangle_{w^m} - \langle x^{m+1}, x^{m+1} \rangle_{w^m}
= \langle x^m + x^{m+1}, x^m - x^{m+1} \rangle_{w^m}
= \langle x^m - x^{m+1}, x^m - x^{m+1} \rangle_{w^m}
= \sum_{j=1}^{N} w_j^m (x_j^m - x_j^{m+1})^2 \geq A^{-1} \|x^m_j - x^{m+1}_j\|_{\ell_2}^2
\]
Key Lemma continued

- The third equality uses the fact that
  \[ \langle x^{m+1}, x^m - x^{m+1} \rangle_{w^m} = 0 \] because of orthogonality.
the third equality uses the fact that
\[ \langle x^{m+1}, x^m - x^{m+1} \rangle_{w^m} = 0 \]
because of orthogonality

sum these inequalities over \( m \geq 1 \)
A related functional

From the monotonicity of $\epsilon_m$, we know that $\epsilon := \lim_{m \to \infty} \epsilon_m$ exists and is non-negative.
A related functional

- From the monotonicity of $\epsilon_m$, we know that $\epsilon := \lim_{m \to \infty} \epsilon_m$ exists and is non-negative.

- $f_\epsilon(z) := \sum_{j=1}^{N} (z_j^2 + \epsilon^2)^{1/2}$
A related functional

- From the monotonicity of $\epsilon_m$, we know that $\epsilon := \lim_{m \to \infty} \epsilon_m$ exists and is non-negative.

- $f_\epsilon(z) := \sum_{j=1}^{N} (z_j^2 + \epsilon^2)^{1/2}$

- $\epsilon > 0$ implies this functional is strictly convex and therefore has a unique minimizer.
A related functional

- From the monotonicity of $\epsilon_m$, we know that $\epsilon := \lim_{m \to \infty} \epsilon_m$ exists and is non-negative

- $f_\epsilon(z) := \sum_{j=1}^{N} (z_j^2 + \epsilon^2)^{1/2}$

- $\epsilon > 0$ implies this functional is strictly convex and therefore has a unique minimizer

- $x_\epsilon := \text{Argmin} \ f_\epsilon(z)$
  \[ z \in \mathcal{F}(y) \]
The minimum of this functional

**Lemma:** Let $\epsilon > 0$ and $\tilde{x} \in \mathcal{F}(y)$. Then $\tilde{x} = x^\epsilon$ if and only if $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}_i = [\tilde{x}_i^2 + \epsilon^2]^{-1/2}$.

For the “only if” part, let $\tilde{x} = x^\epsilon$ and $\eta \in \mathcal{N}$ be arbitrary.
The minimum of this functional

Lemma: Let $\epsilon > 0$ and $\tilde{x} \in \mathcal{F}(y)$. Then $\tilde{x} = x^\epsilon$ if and only if $\langle \tilde{x}, \eta \rangle_{\tilde{w}} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}_i = [\tilde{x}_i^2 + \epsilon^2]^{-1/2}$

For the “only if” part, let $\tilde{x} = x^\epsilon$ and $\eta \in \mathcal{N}$ be arbitrary.

Consider the analytic function $G_\epsilon(t) := f_\epsilon(\tilde{x} + t\eta) - f_\epsilon(\tilde{x})$
The minimum of this functional

Lemma: Let $\epsilon > 0$ and $\tilde{x} \in \mathcal{F}(y)$. Then $\tilde{x} = x^\epsilon$ if and only if $\langle \tilde{x}, \eta \rangle_{\tilde{w}} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}_i = [\tilde{x}_i^2 + \epsilon^2]^{-1/2}$

- For the “only if” part, let $\tilde{x} = x^\epsilon$ and $\eta \in \mathcal{N}$ be arbitrary.
- Consider the analytic function $G_\epsilon(t) := f_\epsilon(\tilde{x} + t\eta) - f_\epsilon(\tilde{x})$
- $G_\epsilon(0) = 0$
Lemma: Let $\epsilon > 0$ and $\tilde{x} \in F(y)$. Then $\tilde{x} = x^\epsilon$ if and only if $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}_i = [\tilde{x}_i^2 + \epsilon^2]^{-1/2}$

- For the “only if” part, let $\tilde{x} = x^\epsilon$ and $\eta \in \mathcal{N}$ be arbitrary.
- Consider the analytic function $G_\epsilon(t) := f_\epsilon(\tilde{x} + t\eta) - f_\epsilon(\tilde{x})$
- $G_\epsilon(0) = 0$
- Hence $G_\epsilon(t) \geq 0$ for all $t \in \mathbb{R}$
The minimum of this functional

**Lemma:** Let $\epsilon > 0$ and $\tilde{x} \in F(y)$. Then $\tilde{x} = x^\epsilon$ if and only if $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}_i = [\tilde{x}^2_i + \epsilon^2]^{-1/2}$

- For the “only if” part, let $\tilde{x} = x^\epsilon$ and $\eta \in \mathcal{N}$ be arbitrary.
- Consider the analytic function $G_\epsilon(t) := f_\epsilon(\tilde{x} + t\eta) - f_\epsilon(\tilde{x})$
- $G_\epsilon(0) = 0$
- Hence $G_\epsilon(t) \geq 0$ for all $t \in \mathbb{R}$
- Hence $G'_\epsilon(0) = 0$
The minimum of this functional

**Lemma:** Let $\epsilon > 0$ and $\tilde{x} \in \mathcal{F}(y)$. Then $\tilde{x} = x^\epsilon$ if and only if $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}_i = [\tilde{x}_i^2 + \epsilon^2]^{-1/2}$

- For the “only if” part, let $\tilde{x} = x^\epsilon$ and $\eta \in \mathcal{N}$ be arbitrary.
- Consider the analytic function $G_\epsilon(t) := f_\epsilon(\tilde{x} + t\eta) - f_\epsilon(\tilde{x})$
- $G_\epsilon(0) = 0$
- Hence $G_\epsilon(t) \geq 0$ for all $t \in IR$
- Hence $G'_\epsilon(0) = 0$
- $G''_\epsilon(0) = \sum_{j=1}^{N} \frac{\eta_i \tilde{x}_i}{[\tilde{x}_i^2 + \epsilon^2]^{1/2}} = \langle \tilde{x}, \eta \rangle \tilde{w}$
The if part of the Proof

For the “if” part, assume that $\tilde{x} \in \mathcal{F}(y)$ and $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}$ as above.
The if part of the Proof

- For the “if” part, assume that $\tilde{x} \in \mathcal{F}(y)$ and $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}$ as above.
- We shall show that $\tilde{x}$ is a minimizer of $f_\epsilon$ on $\mathcal{F}(y)$. 
The if part of the Proof

- For the “if” part, assume that $\tilde{x} \in \mathcal{F}(y)$ and $\langle \tilde{x}, \eta \rangle_{\tilde{w}} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}$ as above.
- We shall show that $\tilde{x}$ is a minimizer of $f_\epsilon$ on $\mathcal{F}(y)$.
- Consider the convex univariate function $[u^2 + \epsilon^2]^{1/2}$.
The if part of the Proof

For the “if” part, assume that $\tilde{x} \in \mathcal{F}(y)$ and $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}$ as above.

We shall show that $\tilde{x}$ is a minimizer of $f_\epsilon$ on $\mathcal{F}(y)$.

consider the convex univariate function $[u^2 + \epsilon^2]^{1/2}$

$[u^2 + \epsilon^2]^{1/2} \geq [u_0^2 + \epsilon^2]^{1/2} + [u_0^2 + \epsilon^2]^{-1/2}u_0(u - u_0)$
The if part of the Proof

For the “if” part, assume that $\tilde{x} \in \mathcal{F}(y)$ and $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}$ as above.

We shall show that $\tilde{x}$ is a minimizer of $f_\epsilon$ on $\mathcal{F}(y)$.

Consider the convex univariate function $[u^2 + \epsilon^2]^{1/2}$

$[u^2 + \epsilon^2]^{1/2} \geq [u_0^2 + \epsilon^2]^{1/2} + [u_0^2 + \epsilon^2]^{-1/2}u_0(u - u_0)$

Linear function tangent to this function at $u_0$
The if part of the Proof

- For the “if” part, assume that $\tilde{x} \in \mathcal{F}(y)$ and $\langle \tilde{x}, \eta \rangle \tilde{w} = 0$ for all $\eta \in \mathcal{N}$, where $\tilde{w}$ as above.

- We shall show that $\tilde{x}$ is a minimizer of $f_\epsilon$ on $\mathcal{F}(y)$.

- Consider the convex univariate function $[u^2 + \epsilon^2]^{1/2}$

- $[u^2 + \epsilon^2]^{1/2} \geq [u_0^2 + \epsilon^2]^{1/2} + [u_0^2 + \epsilon^2]^{-1/2}u_0(u - u_0)$

- Linear function tangent to this function at $u_0$

- Apply this give

\[
f_\epsilon(z) \geq f_\epsilon(\tilde{x}) + \sum_{j=1}^{N} [(\tilde{x}_j)^2 + \epsilon^2]^{-1/2}\tilde{x}_j(z_j - \tilde{x}_j) = f_\epsilon(\tilde{x}) + \langle \tilde{x}, z - \tilde{x} \rangle \tilde{w} = f_\epsilon(\tilde{x})
\]
The if part of the Proof

- For the “if” part, assume that $\tilde{x} \in \mathcal{F}(y)$ and $\langle \tilde{x}, \eta \rangle \bar{w} = 0$ for all $\eta \in \mathcal{N}$, where $\bar{w}$ as above.
- We shall show that $\tilde{x}$ is a minimizer of $f_\epsilon$ on $\mathcal{F}(y)$.
- Consider the convex univariate function $[u^2 + \epsilon^2]^{1/2}$
- $[u^2 + \epsilon^2]^{1/2} \geq [u_0^2 + \epsilon^2]^{1/2} + [u_0^2 + \epsilon^2]^{-1/2}u_0(u - u_0)$
- Linear function tangent to this function at $u_0$
- Apply this give
  
  
  $$f_\epsilon(z) \geq f_\epsilon(\tilde{x}) + \sum_{j=1}^{N}[(\tilde{x}_j)^2 + \epsilon^2]^{-1/2}\tilde{x}_j(z_j - \tilde{x}_j) = f_\epsilon(\tilde{x}) + \langle \tilde{x}, z - \tilde{x} \rangle \bar{w} = f_\epsilon(\tilde{x})$$

- We have used the orthogonality and $z - \tilde{x}$ is in $\mathcal{N}$
The if part of the Proof

- For the “if” part, assume that \( \tilde{x} \in \mathcal{F}(y) \) and \( \langle \tilde{x}, \eta \rangle \tilde{w} = 0 \) for all \( \eta \in \mathcal{N} \), where \( \tilde{w} \) as above.

- We shall show that \( \tilde{x} \) is a minimizer of \( f_\epsilon \) on \( \mathcal{F}(y) \).

- Consider the convex univariate function \( [u^2 + \epsilon^2]^{1/2} \).

- \( [u^2 + \epsilon^2]^{1/2} \geq [u_0^2 + \epsilon^2]^{1/2} + [u_0^2 + \epsilon^2]^{-1/2}u_0(u - u_0) \)

- Linear function tangent to this function at \( u_0 \)

- Apply this give
  \[
  f_\epsilon(z) \geq f_\epsilon(\tilde{x}) + \sum_{j=1}^{N} [(\tilde{x}_j)^2 + \epsilon^2]^{-1/2} \tilde{x}_j(z_j - \tilde{x}_j) = f_\epsilon(\tilde{x}) + \langle \tilde{x}, z - \tilde{x} \rangle \tilde{w} = f_\epsilon(\tilde{x})
  \]

- We have used the orthogonality and \( z - \tilde{x} \) is in \( \mathcal{N} \)

- Since \( z \) is arbitrary, it follows that \( \tilde{x} = x^\epsilon \)
Proof in the Case $\epsilon = 0$

We want to prove that $x^m$ converges to $x^*$
Proof in the Case $\epsilon = 0$

- We want to prove that $x^m$ converges to $x^*$
- Suppose $\epsilon_{n_0} = 0$ for some $n_0$, 
Proof in the Case $\epsilon = 0$

- We want to prove that $x^m$ converges to $x^*$
- Suppose $\epsilon_{n_0} = 0$ for some $n_0$,
- Algorithm stops and outputs $K$ sparse signal which is necessarily $x^*$ - unique sparse solution in $\mathcal{F}(y)$
Proof in the Case $\epsilon = 0$

- We want to prove that $x^m$ converges to $x^*$
- Suppose $\epsilon_{n_0} = 0$ for some $n_0$,
- Algorithm stops and outputs $K$ sparse signal which is necessarily $x^*$ - unique sparse solution in $\mathcal{F}(y)$
- Second case: $\epsilon_m > 0$ for all $m$
Proof in the Case $\epsilon = 0$

- We want to prove that $x^m$ converges to $x^*$
- Suppose $\epsilon_{n_0} = 0$ for some $n_0$,
- Algorithm stops and outputs $K$ sparse signal which is necessarily $x^*$ - unique sparse solution in $\mathcal{F}(y)$
- Second case: $\epsilon_m > 0$ for all $m$
- Since $\epsilon_m \to 0$ there is an increasing sequence of indices $(n_i)$ such that $\epsilon_{n_i} < \epsilon_{n_i-1}$ for all $i$
Proof in the Case $\epsilon = 0$

- We want to prove that $x^m$ converges to $x^*$
- Suppose $\epsilon_{n_0} = 0$ for some $n_0$,
- Algorithm stops and outputs $K$ sparse signal which is necessarily $x^*$ - unique sparse solution in $\mathcal{F}(y)$
- Second case: $\epsilon_m > 0$ for all $m$
- Since $\epsilon_m \to 0$ there is an increasing sequence of indices $(n_i)$ such that $\epsilon_{n_i} < \epsilon_{n_i-1}$ for all $i$
- $r(x^{n_i})_K < N\epsilon_{n_i-1}$ for all $i$
Proof in the Case $\epsilon = 0$

- We want to prove that $x^m$ converges to $x^*$
- Suppose $\epsilon_{n_0} = 0$ for some $n_0$,
- Algorithm stops and outputs $K$ sparse signal which is necessarily $x^*$ - unique sparse solution in $\mathcal{F}(y)$
- Second case: $\epsilon_m > 0$ for all $m$
- Since $\epsilon_m \to 0$ there is an increasing sequence of indices $(n_i)$ such that $\epsilon_{n_i} < \epsilon_{n_{i-1}}$ for all $i$
- $r(x^{n_i})_K < N\epsilon_{n_{i-1}}$ for all $i$
- Since $(x^m)$ is bounded, there is a subsequence $(p_i)$ of $(n_i)$ such that $(x^{p_i})$ converges to a point $\tilde{x} \in \mathcal{F}(y)$
Proof in the Case $\epsilon = 0$

- We want to prove that $x^m$ converges to $x^*$
- Suppose $\epsilon_{n_0} = 0$ for some $n_0$,
- Algorithm stops and outputs $K$ sparse signal which is necessarily $x^*$ - unique sparse solution in $F(y)$
- Second case: $\epsilon_m > 0$ for all $m$
- Since $\epsilon_m \to 0$ there is an increasing sequence of indices $(n_i)$ such that $\epsilon_{n_i} < \epsilon_{n_i-1}$ for all $i$
- $r(x^{n_i})_K < N\epsilon_{n_i-1}$ for all $i$
- Since $(x^m)$ is bounded, there is a subsequence $(p_i)$ of $(n_i)$ such that $(x^{p_i})$ converges to a point $\tilde{x} \in F(y)$
Case $\epsilon = 0$ continued

By Lipschitz continuity $r(x^p_i)_K$ converges to $r(\tilde{x})_K$
Case $\epsilon = 0$ continued

- By Lipschitz continuity $r(x^{p_i})_K$ converges to $r(\tilde{x})_K$
- $r(\tilde{x})_K = \lim_{i \to \infty} r(x^{p_i})_K \leq \lim_{i \to \infty} N \epsilon_{p_{i-1}} = 0$. Hence $\tilde{x}$ has support $K$
Case $\epsilon = 0$ continued

- By Lipschitz continuity $r(x^{p_i})_K$ converges to $r(\tilde{x})_K$
- $r(\tilde{x})_K = \lim_{i \to \infty} r(x^{p_i})_K \leq \lim_{i \to \infty} N \epsilon_{p_i-1} = 0$. Hence $\tilde{x}$ has support $K$
- Since $\Phi$ has RIP property of order $3K$, there is a unique vector in $\mathcal{F}(y)$ with support $\leq K$ and this vector is the minimum $\ell_1$ norm solution
Case $\epsilon = 0$ continued

- By Lipschitz continuity $r(x^{p_i})_K$ converges to $r(\tilde{x})_K$
- $r(\tilde{x})_K = \lim_{i \to \infty} r(x^{p_i})_K \leq \lim_{i \to \infty} N\epsilon_{p_i-1} = 0$. Hence $\tilde{x}$ has support $K$.
- Since $\Phi$ has RIP property of order $3K$, there is a unique vector in $F(y)$ with support $\leq K$ and this vector is the minimum $\ell_1$ norm solution.
- Hence $\tilde{x} = x^*$
Still to be Proved

We must still show that $x^m \rightarrow x^*$
We must still show that $x^m \to x^*$

We know $x^p_i \to x^*$ and $\epsilon p_i \to 0$
Still to be Proved

- We must still show that $x^m \rightarrow x^*$
- We know $x^{p_i} \rightarrow x^*$ and $\epsilon_{p_i} \rightarrow 0$
- Hence $\mathcal{J}(x^{p_i}, w^{p_i}, \epsilon_{p_i}) \rightarrow \|x^*\|_{\ell_1}$
Still to be Proved

- We must still show that $x^m \rightarrow x^*$
- We know $x^{p_i} \rightarrow x^*$ and $\epsilon_{p_i} \rightarrow 0$
- Hence $J(x^{p_i}, w^{p_i}, \epsilon_{p_i}) \rightarrow \|x^*\|_1$
- Monotonicity property implies $J(x^m, w^m, \epsilon_m) \rightarrow \|x^*\|_1$
We must still show that \( x^m \rightarrow x^* \).

We know \( x^{p_i} \rightarrow x^* \) and \( \epsilon_{p_i} \rightarrow 0 \).

Hence \( \mathcal{J}(x^{p_i}, w^{p_i}, \epsilon_{p_i}) \rightarrow \|x^*\|_1 \).

monotonicity property implies \( \mathcal{J}(x^m, w^m, \epsilon_m) \rightarrow \|x^*\|_1 \).

\( \mathcal{J}(x^m, w^m, \epsilon_m) - N\epsilon_m \leq \|x^m\|_1 \leq \mathcal{J}(x^m, w^m, \epsilon_m) \).
Still to be Proved

- We must still show that $x^m \to x^*$
- We know $x^{p_i} \to x^*$ and $\epsilon_{p_i} \to 0$
- Hence $\mathcal{J}(x^{p_i}, w^{p_i}, \epsilon_{p_i}) \to \|x^*\|_{\ell_1}$
- Monotonicity property implies $\mathcal{J}(x^m, w^m, \epsilon_m) \to \|x^*\|_{\ell_1}$
- $\mathcal{J}(x^m, w^m, \epsilon_m) - N\epsilon_m \leq \|x^m\|_{\ell_1} \leq \mathcal{J}(x^m, w^m, \epsilon_m)$
- Hence $\|x^m\|_{\ell_1} \to \|x^*\|_{\ell_1}$
Still to be Proved

- We must still show that $x^m \rightarrow x^*$
- We know $x^{p_i} \rightarrow x^*$ and $\epsilon_{p_i} \rightarrow 0$
- Hence $\mathcal{J}(x^{p_i}, w^{p_i}, \epsilon_{p_i}) \rightarrow \|x^*\|_{\ell_1}$
- Monotonicity property implies $\mathcal{J}(x^m, w^m, \epsilon_m) \rightarrow \|x^*\|_{\ell_1}$
- $\mathcal{J}(x^m, w^m, \epsilon_m) - N\epsilon_m \leq \|x^m\|_{\ell_1} \leq \mathcal{J}(x^m, w^m, \epsilon_m)$
- Hence $\|x^m\|_{\ell_1} \rightarrow \|x^*\|_{\ell_1}$
- Finally, the geometry lemma says

$$\limsup_{m \rightarrow \infty} \|x^m - x^*\|_{\ell_1} \leq \frac{1 + \gamma}{1 - \gamma} \left( \lim_{m \rightarrow \infty} \|x^m\|_{\ell_1} - \|x^*\|_{\ell_1} \right) \leq 0$$
We must still show that \( x^m \to x^* \)

We know \( x^{pi} \to x^* \) and \( \epsilon_{pi} \to 0 \)

Hence \( J(x^{pi}, w^{pi}, \epsilon_{pi}) \to \|x^*\|_1 \)

Monotonicity property implies \( J(x^m, w^m, \epsilon_m) \to \|x^*\|_1 \)

\( J(x^m, w^m, \epsilon_m) - N\epsilon_m \leq \|x^m\|_1 \leq J(x^m, w^m, \epsilon_m) \)

Hence \( \|x^m\|_1 \to \|x^*\|_1 \)

Finally, the geometry lemma says

\[
\limsup_{m \to \infty} \|x^m - x^*\|_1 \leq \frac{1 + \gamma}{1 - \gamma} \left( \lim_{m \to \infty} \|x^m\|_1 - \|x^*\|_1 \right) \leq 0
\]

\( x^m \to x^* \).
We must still show that $x^m \to x^*$

We know $x^{p_i} \to x^*$ and $\epsilon_{p_i} \to 0$

Hence $\mathcal{J}(x^{p_i}, w^{p_i}, \epsilon_{p_i}) \to \|x^*\|_{\ell_1}$

monotonicity property implies $\mathcal{J}(x^m, w^m, \epsilon_m) \to \|x^*\|_{\ell_1}$

$\mathcal{J}(x^m, w^m, \epsilon_m) - N\epsilon_m \leq \|x^m\|_{\ell_1} \leq \mathcal{J}(x^m, w^m, \epsilon_m)$

Hence $\|x^m\|_{\ell_1} \to \|x^*\|_{\ell_1}$

Finally, the geometry lemma says

$$\limsup_{m \to \infty} \|x^m - x^*\|_{\ell_1} \leq \frac{1 + \gamma}{1 - \gamma} \left( \lim_{m \to \infty} \|x^m\|_{\ell_1} - \|x^*\|_{\ell_1} \right)$$

$x^m \to x^*$.
Proof in the Case $\epsilon > 0$

We first show that $x^m \to x^\epsilon$, $n \to \infty$, with $x^\epsilon$ the minimizer of $f$. 

IMA, 2007 – p. 17/23
Proof in the Case $\epsilon > 0$

- We first show that $x^m \to x^\epsilon$, $n \to \infty$, with $x^\epsilon$ the minimizer of $f$.

- Let $(x^{ni})$ be any convergent subsequence of $(x^m)$ and let $\bar{x} \in \mathcal{F}(y)$ be its limit. We want to show that $\bar{x} = x^\epsilon$. 
Proof in the Case $\epsilon > 0$

- We first show that $x^m \to x^\epsilon$, $n \to \infty$, with $x^\epsilon$ the minimizer of $f$
- Let $(x^{ni})$ be any convergent subsequence of $(x^m)$ and let $\tilde{x} \in \mathcal{F}(y)$ be its limit. We want to show that $\tilde{x} = x^\epsilon$
- Since $w_j^n = [(x_j^n)^2 + \epsilon_n^2]^{-1/2} \leq \epsilon^{-1}$
Proof in the Case $\epsilon > 0$

- We first show that $x^m \to x^\epsilon$, $n \to \infty$, with $x^\epsilon$ the minimizer of $f$

- Let $(x^{ni})$ be any convergent subsequence of $(x^m)$ and let $\tilde{x} \in F(y)$ be its limit. We want to show that $\tilde{x} = x^\epsilon$

- Since $w_j^n = [(x_j^n)^2 + \epsilon_n^2]^{-1/2} \leq \epsilon^{-1}$

- $\lim_{i \to \infty} w_j^{ni} = [(\tilde{x}_j)^2 + \epsilon^2]^{-1/2} =: \tilde{w}_j$, $j = 1, \ldots, N$
Proof in the Case $\epsilon > 0$

- We first show that $x^m \to x^\epsilon$, $n \to \infty$, with $x^\epsilon$ the minimizer of $f$

- Let $(x^{ni})$ be any convergent subsequence of $(x^m)$ and let $\tilde{x} \in \mathcal{F}(y)$ be its limit. We want to show that $\tilde{x} = x^\epsilon$

- Since $w^m_j = [(x^m_j)^2 + \epsilon^2_n]^{-1/2} \leq \epsilon^{-1}$

- $\lim_{i \to \infty} w^{ni}_j = [(	ilde{x}_j)^2 + \epsilon^2]^{-1/2} =: \tilde{w}_j, j = 1, \ldots, N$

- Also know $x^{ni+1} \to \tilde{x}$, $i \to \infty$
Proof in the Case $\epsilon > 0$

- We first show that $x^m \to x^\epsilon$, $n \to \infty$, with $x^\epsilon$ the minimizer of $f$

- Let $(x^{ni})$ be any convergent subsequence of $(x^m)$ and let $\tilde{x} \in \mathcal{F}(y)$ be its limit. We want to show that $\tilde{x} = x^\epsilon$

- Since $w_j^n = [(x_j^n)^2 + \epsilon_n^2]^{-1/2} \leq \epsilon^{-1}$

- $\lim_{i \to \infty} w_j^{ni} = [(\tilde{x}_j)^2 + \epsilon^2]^{-1/2} =: \tilde{w}_j, j = 1, \ldots, N$

- Also know $x^{ni+1} \to \tilde{x}, i \to \infty$

- Orthogonality says $\langle \tilde{x}, \eta \rangle_{\tilde{w}} = \lim_{i \to \infty} \langle \tilde{x}^{ni+1}, \eta \rangle_{\tilde{w}^{ni}} = 0$. 
Proof in the Case $\epsilon > 0$

- We first show that $x^m \to x^\epsilon$, $n \to \infty$, with $x^\epsilon$ the minimizer of $f$.

- Let $(x^{ni})$ be any convergent subsequence of $(x^m)$ and let $\tilde{x} \in \mathcal{F}(y)$ be its limit. We want to show that $\tilde{x} = x^\epsilon$.

- Since $w^m_j = [(x^m_j)^2 + \epsilon_n^2]^{-1/2} \leq \epsilon^{-1}$

- $\lim_{i \to \infty} w^{ni}_j = [(\tilde{x}_j)^2 + \epsilon^2]^{-1/2} =: \tilde{w}_j$, $j = 1, \ldots, N$.

- Also know $x^{ni+1} \to \tilde{x}$, $i \to \infty$.

- Orthogonality says $\langle \tilde{x}, \eta \rangle \tilde{w} = \lim_{i \to \infty} \langle \tilde{x}^{ni+1}, \eta \rangle \tilde{w}^{ni} = 0$.

- The “if” part of Lemma for $f_\epsilon$ implies that $\tilde{x} = x^\epsilon$. 

IMA, 2007 – p. 17/23
Proof in the Case $\epsilon > 0$

- We first show that $x^m \to x^\epsilon$, $n \to \infty$, with $x^\epsilon$ the minimizer of $f$

- Let $(x^{n_i})$ be any convergent subsequence of $(x^m)$ and let $\tilde{x} \in \mathcal{F}(y)$ be its limit. We want to show that $\tilde{x} = x^\epsilon$

- Since $w^m_j = [(x^m_j)^2 + \epsilon^2]^{-1/2} \leq \epsilon^{-1}$

- $\lim_{i \to \infty} w^{n_i}_j = [((\tilde{x}_j)^2 + \epsilon^2]^{-1/2} =: \tilde{w}_j, j = 1, \ldots, N$

- Also know $x^{n_i+1} \to \tilde{x}, i \to \infty$

- Orthogonality says $\langle \tilde{x}, \eta \rangle \tilde{w} = \lim_{i \to \infty} \langle \tilde{x}^{n_i+1}, \eta \rangle \tilde{w}^{n_i} = 0$.

- The “if” part of Lemma for $f_\epsilon$ implies that $\tilde{x} = x^\epsilon$.

- This shows $\lim_{m \to \infty} x^m = \tilde{x}$
Proof of error estimate

To prove the error estimate, we first note that
\[ \| x^e \|_{\ell_1} \leq f_\epsilon(x^e) \leq f_\epsilon(x^*) \leq \| x^* \|_{\ell_1} + N \epsilon \]
Proof of error estimate

To prove the error estimate, we first note that
\[ \|x^e\|_{\ell_1} \leq f_\epsilon(x^e) \leq f_\epsilon(x^*) \leq \|x^*\|_{\ell_1} + N\epsilon \]

The second inequality is the minimizing property of $x^e$. 
Proof of error estimate

To prove the error estimate, we first note that
\[ \|x^e\|_{\ell_1} \leq f_\epsilon(x^e) \leq f_\epsilon(x^*) \leq \|x^*\|_{\ell_1} + N\epsilon \]

The second inequality is the minimizing property of \(x^e\)

The third inequality is \(\sqrt{u^2 + \epsilon^2} \leq |u| + \epsilon\)
Proof of error estimate

- To prove the error estimate, we first note that
  \[\|x^\epsilon\|_{\ell_1} \leq f_\epsilon(x^\epsilon) \leq f_\epsilon(x^*) \leq \|x^*\|_{\ell_1} + N\epsilon\]
- The second inequality is the minimizing property of \(x^\epsilon\)
- The third inequality is \(\sqrt{u^2 + \epsilon^2} \leq |u| + \epsilon\)
- Apply the geometrical lemma to obtain
  \[\|x^\epsilon - x^*\|_{\ell_1} \leq \frac{1+\gamma}{1-\gamma}[N\epsilon + 2\sigma_k(x^*)_{\ell_1}]\]
Proof of error estimate

To prove the error estimate, we first note that
\[ \|x^\epsilon\|_{\ell_1} \leq f_\epsilon(x^\epsilon) \leq f_\epsilon(x^*) \leq \|x^*\|_{\ell_1} + N\epsilon \]

The second inequality is the minimizing property of \(x^\epsilon\).

The third inequality is \(\sqrt{u^2 + \epsilon^2} \leq |u| + \epsilon\).

Apply the geometrical lemma to obtain
\[ \|x^\epsilon - x^*\|_{\ell_1} \leq \frac{1+\gamma}{1-\gamma} \left[ N\epsilon + 2\sigma_k(x^*)_{\ell_1} \right] \]

Since \(N\epsilon = \lim_{m \to \infty} N\epsilon_m \leq \lim_{m \to \infty} r(x^m)_K = r(x^\epsilon)_K \)
Proof of error estimate

To prove the error estimate, we first note that
\[ \|x^\epsilon\|_{\ell_1} \leq f_\epsilon(x^\epsilon) \leq f_\epsilon(x^*) \leq \|x^*\|_{\ell_1} + N\epsilon \]
The second inequality is the minimizing property of \(x^\epsilon\)
The third inequality is \(\sqrt{u^2 + \epsilon^2} \leq |u| + \epsilon\)

Apply the geometrical lemma to obtain
\[ \|x^\epsilon - x^*\|_{\ell_1} \leq \frac{1+\gamma}{1-\gamma}[N\epsilon + 2\sigma_k(x^*)_{\ell_1}] \]

Since \(N\epsilon = \lim_{m \to \infty} N\epsilon_m \leq \lim_{m \to \infty} r(x^m)_K = r(x^\epsilon)_K\)

Properties of rearrangements gives
\[
(K - k)N\epsilon \leq (K - k)r(x^\epsilon)_K \\
\leq \|x^\epsilon - x^*\|_{\ell_1} + \sigma_k(x^*)_{\ell_1} \\
\leq \frac{1+\gamma}{1-\gamma}[N\epsilon + 2\sigma_k(x^*)_{\ell_1}] + \sigma_k(x^*)_{\ell_1}
\]
Final Part of Proof

By assumption on $K$, we have $K - k \geq 6 \geq 2\frac{1+\gamma}{1-\gamma}$
Final Part of Proof

By assumption on $K$, we have $K - k \geq 6 \geq 2^{\frac{1+\gamma}{1-\gamma}}$

Hence $N \epsilon \leq 3\sigma_k(x)_{\ell_1}$.
Exponential Convergence

We shall now prove the exponential convergence theorem:

**Theorem** For a given $0 < \rho < 1$, assume $\Phi$ satisfies NSP of order $3K$ with constant $\gamma$ such that

$$\mu := \frac{\gamma}{1 - \rho} \left(1 + \frac{1}{K-k}\right) < 1.$$ Let $m_0$ be such that

$$\|x^{m_0} - x^*\|_{\ell_1} \leq R^* := \rho \min_{i \in T} |x_i| = \rho r(x)_k.$$ Then for all $m \geq m_0$, we have

$$\|x^{m+1} - x^*\|_{\ell_1} \leq \mu \|x^m - x^*\|_{\ell_1}.$$ Consequently $x^m$ converges to $x^*$ exponentially.
Proof of exponential convergence

Suppose $x^*$ is $k$ sparse
Proof of exponential convergence

- Suppose $x^*$ is $k$ sparse
- $\eta^m \in \mathcal{N}$ such that $\eta^m := x^m - x^*$. 
Proof of exponential convergence

- Suppose $x^*$ is $k$ sparse
- $\eta^m \in \mathcal{N}$ such that $\eta^m := x^m - x^*$.
- $E_m := \|\eta^m\|_{\ell_1}$
Proof of exponential convergence

- Suppose $x^*$ is $k$ sparse
- $\eta^m \in \mathcal{N}$ such that $\eta^m := x^m - x^*$.
- $E_m := \|\eta^m\|_{\ell_1}$
- We know that $E_m \to 0$
Proof of exponential convergence

- Suppose $x^*$ is $k$ sparse
- $\eta^m \in \mathcal{N}$ such that $\eta^m := x^m - x^*$.
- $E_m := \|\eta^m\|_{\ell_1}$
- We know that $E_m \to 0$
- Orthogonality gives $\sum_i (x_i^* + \eta_i^{m+1})\eta_i^{m+1} w_i^m = 0$
Proof of exponential convergence

Suppose \( x^* \) is \( k \) sparse

\[ \eta^m \in \mathcal{N} \] such that \( \eta^m := x^m - x^* \).

\[ E_m := \| \eta^m \|_{\ell_1} \]

We know that \( E_m \to 0 \)

Orthogonality gives \( \sum_i (x^*_i + \eta^*_i) \eta^{m+1}_i w^m_i = 0 \)

Hence

\[
\sum_i |\eta^{m+1}_i|^2 w^m_i = - \sum_i x^*_i \eta^{m+1}_i w^m_i
\]

\[
= - \sum_{i \in T} \frac{x^*_i}{[(x^*_i)^2 + \epsilon^2_i]^1/2} \eta^{m+1}_i
\]
Proof of exponential convergence

- Suppose $x^*$ is $k$-sparse
- $\eta^m \in \mathcal{N}$ such that $\eta^m := x^m - x^*$.
- $E_m := \|\eta^m\|_1$
- We know that $E_m \to 0$
- Orthogonality gives $\sum_i (x^*_i + \eta^{m+1}_i)\eta^{m+1}_i w^m_i = 0$
- Hence
  \[
  \sum_i |\eta^{m+1}_i|^2 w^m_i = - \sum_i x^*_i \eta^{m+1}_i w^m_i \\
  = - \sum_{i \in T} \frac{x^*_i}{[(x^*_i)^2 + \epsilon^2_m]^{1/2}} \eta^{m+1}_i 
  \]

- Assume $E_m \leq R^*$
The Proof continued

\[ |\eta^m_i| \leq \|\eta^m\|_{\ell_1} \leq \rho |x^*_i|, \quad i \in T \]
The Proof continued

- \(|\eta_i^m| \leq \|\eta^m\|_{\ell_1} \leq \rho|x_i^*|, \quad i \in T\)

- Hence \(\frac{|x_i^*|}{[(x_i^m)^2 + \epsilon^2_m]^{1/2}} \leq |x_i^* + \eta_i^m| \leq \frac{1}{1-\rho}\)
The Proof continued

- \(|\eta_i^m| \leq \|\eta^m\|_{\ell_1} \leq \rho|x_i^*|, \quad i \in T\)
- Hence \(\frac{|x_i^*|}{[(x_i^m)^2 + \epsilon_m^2]^{1/2}} \leq \frac{|x_i^*|}{|x_i^* + \eta_i^m|} \leq \frac{1}{1-\rho}\)
- NSP gives \(\sum_i |\eta_i^{m+1}|^2 w_i^m \leq \frac{1}{1-\rho} \|\eta_T^{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} \|\eta_T^{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} \|\eta^{m+1}\|_{\ell_1}\)
The Proof continued

- \(|\eta_i^m| \leq \|\eta^m\| \leq \rho|x_i^*|, \quad i \in T\)

- Hence
  \[
  \frac{|x_i^*|}{[(x_i^m)^2 + \epsilon_m^2]^{1/2}} \leq \frac{|x_i^*|}{|x_i^* + \eta_i^m|} \leq \frac{1}{1 - \rho}
  \]

- NSP gives
  \[
  \sum_i |\eta_i^{m+1}|^2 w_i^m \leq \frac{1}{1 - \rho} \|\eta_T^{m+1}\| \leq \frac{\gamma}{1 - \rho} \|\eta_T^{m+1}\| \leq \frac{\gamma}{1 - \rho} \|\eta^{m+1}\| \leq \gamma \|\eta^{m+1}\| \leq \gamma \|\eta^{m+1}\| \leq \gamma \|\eta^{m+1}\|
  \]

- Hence (writing \(\eta_i = w_i^{1/2} \eta_i w_i^{-1/2}\)) from Cauchy-Schwarz
  \[
  \|\eta^{m+1}\| \leq \left( \sum |\eta_i^{m+1}|^2 w_i^m \right) \left( \sum [(\eta_i^m)^2 + \epsilon_m^2]^{1/2} \right) \leq \frac{\gamma}{1 - \rho} \|\eta^{m+1}\| \leq \gamma \|\eta^{m+1}\| \leq \gamma \|\eta^{m+1}\| \leq \gamma \|\eta^{m+1}\| \leq \gamma \|\eta^{m+1}\| \leq \gamma \|\eta^{m+1}\|
  \]
Final Touches

- If $\eta^{m+1} = 0$, then $x^{m+1} = x^*$
Final Touches

- If $\eta^{m+1} = 0$, then $x^{m+1} = x^*$
- Otherwise $\|\eta^{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} (\|\eta^m\|_{\ell_1} + N\epsilon_m)$
If $\eta^{m+1} = 0$, then $x^{m+1} = x^*$

Otherwise $\|\eta^{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} (\|\eta^m\|_{\ell_1} + N\epsilon_m)$

However

$$N\epsilon_m \leq r(x^n)K \leq \frac{1}{K-k}(\|x^m - x^*\|_{\ell_1} + \sigma_k(x^*)_{\ell_1}) = \frac{\|\eta^m\|_{\ell_1}}{K-k}$$
If \( \eta^{m+1} = 0 \), then \( x^{m+1} = x^* \)

Otherwise \( \|\eta^{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} (\|\eta^m\|_{\ell_1} + N\epsilon_m) \)

However \( N\epsilon_m \leq r(x^n)K \leq \frac{1}{K-k} (\|x^m - x^*\|_{\ell_1} + \sigma_k(x^*)_{\ell_1}) = \frac{\|\eta^m\|_{\ell_1}}{K-k} \)

Finally

\[ E_{m+1} = \|\eta^{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} \left( 1 + \frac{1}{K-k} \right) \|\eta^m\|_{\ell_1} = \mu E_m. \]
Final Touches

- If $\eta_{m+1} = 0$, then $x_{m+1} = x^*$

- Otherwise $\|\eta_{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} (\|\eta_{m}\|_{\ell_1} + N\epsilon_m)$

- However $N\epsilon_m \leq r(x^n) K \leq \frac{1}{K-k} (\|x_m - x^*\|_{\ell_1} + \sigma_k(x^*)_{\ell_1}) = \frac{\|\eta_{m}\|_{\ell_1}}{K-k}$

- Finally

$$E_{m+1} = \|\eta_{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} \left(1 + \frac{1}{K-k}\right) \|\eta_{m}\|_{\ell_1} = \mu E_m.$$ 

- Since $\mu < 1$, we can increment our induction hypothesis to get $E_{m+1} \leq R^*$
Final Touches

- If $\eta^{m+1} = 0$, then $x^{m+1} = x^*$
- Otherwise $\|\eta^{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} (\|\eta^m\|_{\ell_1} + N\epsilon_m)$
- However
  \[ N\epsilon_m \leq r(x^n)K \leq \frac{1}{K-k}(\|x^m - x^*\|_{\ell_1} + \sigma_k(x^*)_{\ell_1}) = \frac{\|\eta^m\|_{\ell_1}}{K-k} \]
- Finally
  \[ E_{m+1} = \|\eta^{m+1}\|_{\ell_1} \leq \frac{\gamma}{1-\rho} \left(1 + \frac{1}{K-k}\right) \|\eta^m\|_{\ell_1} = \mu E_m. \]
- Since $\mu < 1$, we can increment our induction hypothesis to get $E_{m+1} \leq R^*$