Selected recipes for inverse problems

Ferréol Soulez

Notations

α	(lower case) scalar
α^*	complex conjugate of α
$\boldsymbol{x} \in \mathbb{C}^N$	(boldface lower case) complex vector
N	(upper case) cardinal (number of elements)
$\mathbf{H} \in \mathbb{C}^{N\times M}$	(boldface upper case) linear operator (matrix)
$f : \mathbb{C}^N \to \mathbb{C}^M$	(lower case) function
$L^2(\mathbb{R}^N)$	space of squared-integrable function
$\mathcal{H} : L^2(\mathbb{R}^N) \to L^2(\mathbb{R}^M)$	(calligraphic upper case) operator acting on functions

Basic properties

$[\mathbf{H}\,\boldsymbol{x}]_m = \sum_n H_{m,n}\, x_n$ — vector matrix product

$[\mathbf{H}\,\mathbf{G}]_{m,k} = \sum_n H_{m,n}\, G_{n,k}$ — matrix product

$\mathbf{H}\,(\alpha\,\boldsymbol{x} + \beta\,\boldsymbol{y}) = \alpha\,\mathbf{H}\,\boldsymbol{x} + \beta\,\mathbf{H}\,\boldsymbol{y}$ — linearity

$(\boldsymbol{x} \times \boldsymbol{y})_n = x_n\, y_n$ — element wise product

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$ identity matrix

$$\mathbf{A} = \mathrm{diag}(\boldsymbol{a}) = \begin{bmatrix} a_0 & 0 & 0 & \dots & 0 \\ 0 & a_1 & 0 & \dots & 0 \\ 0 & 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{N-1} \end{bmatrix}$$ diagonal matrix

Adjoint, scalar product

$\left(H^\dagger\right)_{i,j} = H^*_{j,i}$ — adjoint (transpose conjugate)

$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{n=0}^{N-1} x_n^*\, y_n = \boldsymbol{x}^\dagger \boldsymbol{y}$ — dot (or scalar or inner) product

$\langle \mathbf{H}\,\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \mathbf{H}^\dagger\,\boldsymbol{y} \rangle$ — adjoint formal definition

$(\alpha\,\mathbf{H} + \beta\,\mathbf{G})^\dagger = \alpha^*\,\mathbf{H}^\dagger + \beta^*\,\mathbf{G}^\dagger$

$(\mathbf{H}\,\mathbf{G})^\dagger = \mathbf{G}^\dagger\,\mathbf{H}^\dagger$

Norms

$\|\boldsymbol{x}\|_2^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{x}^\dagger \boldsymbol{x} = \sum_{n=0}^{N-1} |x_n|^2 = \sum_{n=0}^{N-1} x_n^*\, x_n$ — ℓ^2 norm

$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\|_2\, \|\boldsymbol{y}\|_2$ — (Cauchy-Schwartz)

$\|\boldsymbol{x} + \boldsymbol{y}\|_2^2 = \|\boldsymbol{x}\|_2^2 + \|\boldsymbol{y}\|_2^2 + 2\,\langle \boldsymbol{x}, \boldsymbol{y} \rangle$

$\|\boldsymbol{x}\|_p^p = \sum_{n=0}^{N-1} |x_n|^p \ ,\ p \geq 1$ — ℓ^p norm

$\|\boldsymbol{x} + \boldsymbol{y}\|_p \leq \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p$ — triangular inequality

$\|\mathbf{H}\|_F^2 = \sum_{n,m} |H_{i,j}|^2 = \mathrm{tr}(\mathbf{H}^\dagger \mathbf{H})$ — Frobenius norm

Inverse

$\mathbf{H} \in \mathbb{C}^{N\times N}$ is called invertible (also nonsingular or nondegenerate) if there exists a matrix $\mathbf{G}$ such that:

$$\mathbf{HG} = \mathbf{GH} = \mathbf{I}$$

$\mathbf{G} = \mathbf{H}^{-1}$ is unique and is the inverse of $\mathbf{H}$

$(\mathbf{H}\,\mathbf{M})^{-1} = \mathbf{M}^{-1}\,\mathbf{H}^{-1}$ — for any $\mathbf{H}\,,\mathbf{M}$ invertible

$\left(\mathbf{H}^\dagger\right)^{-1} = \left(\mathbf{H}^{-1}\right)^\dagger$

Eigenvalue & eigenvectors of $\mathbf{H} \in \mathbb{C}^{N\times N}$

$\lambda_i \in \mathbb{C}$ and $\boldsymbol{v}_i \in \mathbb{C}^N$ are the i^{th} eigenvalue and eigenvector respectively. They satisfy:

$$\mathbf{H}\,\boldsymbol{v}_i = \lambda_i\,\boldsymbol{v}_i$$

It leads to the eigendecomposition:

$$\mathbf{H} = \mathbf{Q}\ \mathrm{diag}(\boldsymbol{\lambda})\,\mathbf{Q}^{-1}\,,$$

where the columns of $\mathbf{V}$ are the N eigenvectors of $\mathbf{H}$.

- $\boldsymbol{\lambda} = \mathrm{eig}(\mathbf{H})$ is the spectrum of $\mathbf{H}$,
- $C = \frac{\max|\boldsymbol{\lambda}|}{\min|\boldsymbol{\lambda}|}$ is the condition number,
- $\mathrm{rank}(\mathbf{H})$ is the number of non-zero element of $\boldsymbol{\lambda}$,
- if $\lambda_i \neq 0$, $\forall i$, $\mathbf{H}$ is invertible and $\mathbf{H}^{-1} = \mathbf{Q}\ \mathrm{diag}(\boldsymbol{\lambda})^{-1}\,\mathbf{Q}^{-1}$,

Peculiar matrices

- **Unitary:** $\mathbf{Q}^\dagger\,\mathbf{Q} = \mathbf{Q}\,\mathbf{Q}^\dagger = \mathbf{I}$
- **Hermitian:** $\mathbf{H}^\dagger = \mathbf{H}$ and $\mathbf{H} = \mathbf{Q}\ \mathrm{diag}(\boldsymbol{\lambda})\,\mathbf{Q}^\dagger$ with $\boldsymbol{\lambda} \in \mathbb{R}^N$ and $\mathbf{Q}$ unitary
- **Positive semi-definite:** $\lambda_i \geq 0\,,\ \forall i \iff \boldsymbol{x}^\dagger \mathbf{H} \boldsymbol{x} \geq 0,\ \forall \boldsymbol{x}$
- **Toeplitz:** $H_{i,j} = h_{i-j}$,
- **Circulant:** Toeplitz with $H_{i,j} = h_{(i-j) \mod N}$

Trace and determinant

- **trace:** $\mathrm{tr}(\mathbf{H}) = \sum_n H_{n,n} = \sum_n \lambda_n$

$\mathrm{tr}(\mathbf{H}^\dagger) = \mathrm{tr}(\mathbf{H})^*$ $\qquad$ $\mathrm{tr}(\mathbf{H}\,\mathbf{G}) = \mathrm{tr}(\mathbf{G}\,\mathbf{H})$

$\mathrm{tr}(\alpha\,\mathbf{H} + \beta\,\mathbf{G}) = \alpha\ \mathrm{tr}(\mathbf{H}) + \alpha\ \mathrm{tr}(\mathbf{H})$

- **determinant:** $\det(\mathbf{H}) = \prod_n \lambda_n$

$\det(\mathbf{H}) \neq 0 \iff \mathbf{H}$ is invertible $\qquad$ $\det(\mathbf{H}^\dagger) = \det(\mathbf{H})^*$

$\det(\mathbf{H}\,\mathbf{G}) = \det(\mathbf{H})\ \det(\mathbf{G})$ $\qquad$ $\det(\mathbf{H}^{-1}) = 1/\det(\mathbf{H})$

Singular value decomposition

For all matrices $\mathbf{H} \in \mathbb{C}^{N\times M}$ there exists two unitary matrices $\mathbf{U} \in \mathbb{C}^{N\times N}$ and $\mathbf{V} \in \mathbb{C}^{M\times M}$ and a real non-negative diagonal matrix $\boldsymbol{\Sigma} \in \mathbb{C}^{N\times M}$ (singular values: $\sigma_i = \Sigma_{i,i}\,,\ \forall i \leq \min(M,N)$) subject to:

$$\mathbf{H} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^\dagger$$

- columns of $\mathbf{U}$ are eigenvectors of $\mathbf{H}\,\mathbf{H}^\dagger$,
- columns of $\mathbf{V}$ are eigenvectors of $\mathbf{H}^\dagger\,\mathbf{H}$,
- $\boldsymbol{\Sigma} = \sqrt{\mathrm{diag}\left(\mathrm{eig}\left(\mathbf{H}\,\mathbf{H}^\dagger\right)\right)}$,
- $\|\mathbf{H}\|_F^2 = \sum_i \sigma_i^2$
- $\mathrm{rank}(\mathbf{H})$: number of non zero singular values,

Inversion lemmas & Woodbury identity

$$\mathbf{B}^{-1}\mathbf{V}\left(\mathbf{A} - \mathbf{U}\mathbf{B}^{-1}\mathbf{V}\right)^{-1} = \left(\mathbf{B} - \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{A}^{-1}$$

$$\left(\mathbf{A} - \mathbf{U}\mathbf{B}^{-1}\mathbf{V}\right)^{-1}\mathbf{U}\mathbf{B}^{-1} = \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{B} - \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}$$

$$\left(\mathbf{A} + \mathbf{U}\,\mathbf{B}\,\mathbf{V}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\,\mathbf{U}\left(\mathbf{B}^{-1} + \mathbf{V}\,\mathbf{A}^{-1}\,\mathbf{U}\right)^{-1}\,\mathbf{V}\,\mathbf{A}^{-1}$$

Moore-Penrose pseudo inverse

The pseudo inverse of $\mathbf{H} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^\dagger$ writes $\mathbf{H}^+ = \mathbf{V}\,\boldsymbol{\Sigma}^+\,\mathbf{U}^\dagger$ with $\Sigma^+_{i,i} = \begin{cases} \Sigma^{-1}_{i,i} & \text{if } \sigma_i \neq 0\,, \\ 0 & \text{otherwise}\,. \end{cases}$

- $\mathbf{H}\,\mathbf{H}^+\,\mathbf{H} = \mathbf{H}$ and $\mathbf{H}^+\,\mathbf{H}\,\mathbf{H}^+ = \mathbf{H}^+$
- $\mathbf{H}$ is square and $\mathrm{rank}(\mathbf{H}) = N \Rightarrow \mathbf{H}^+ = \mathbf{H}^{-1}$
- $\mathbf{H}$ is broad: $\mathrm{rank}(\mathbf{H}) \leq N$ and $\mathbf{H}^+ = \mathbf{H}^\dagger\left(\mathbf{H}\,\mathbf{H}^\dagger\right)^{-1}$
- $\mathbf{H}$ is tall: $\mathrm{rank}(\mathbf{H}) \leq M$ and $\mathbf{H}^+ = \left(\mathbf{H}^\dagger\,\mathbf{H}\right)^{-1}\,\mathbf{H}^\dagger$

Selected recipes for inverse problems

Ferréol Soulez

Derivatives

$J_{i,j} = \left[\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right]_{i,j} = \frac{\partial x_i}{\partial y_j}$ — Jacobian matrix

$[\nabla f]_i = \frac{\partial f}{\partial x_i}$ — Gradient vector

$[\nabla^2 f]_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ — Hessian matrix

Derivation rules

$\partial\,(\alpha\,\mathbf{A} + \beta\,\mathbf{B}) = \alpha\,\partial\mathbf{A} + \beta\,\partial\mathbf{B}$ — linearity

$\partial\,(\mathbf{A}\,\mathbf{B}) = (\partial\mathbf{A})\ \mathbf{B} + \mathbf{A}\ (\partial\mathbf{B})$

$\partial\,(\mathbf{A}^{-1}) = -\mathbf{A}^{-1}\,(\partial\mathbf{A})\ \mathbf{A}^{-1}$

$\partial\left(\mathbf{A}^{\dagger}\right) = (\partial\mathbf{A})^{\dagger}$

$\frac{\partial \boldsymbol{x}^{\dagger}\boldsymbol{y}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{y}^{\dagger}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{y}$

$\frac{\partial \boldsymbol{x}^{\dagger}\mathbf{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \left(\mathbf{A} + \mathbf{A}^{\dagger}\right)\boldsymbol{x}$

$\frac{\partial f\circ g(\boldsymbol{x})}{\partial \boldsymbol{x}} = \sum_{m=0}^{M-1} \frac{\partial f}{\partial u_m}\frac{\partial u_m}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{u}=g(\boldsymbol{x})}$ — chain rule

$\frac{\partial (\mathbf{A}\,\boldsymbol{x}-\boldsymbol{y})^{\dagger}\mathbf{B}(\mathbf{A}\,\boldsymbol{x}-\boldsymbol{y})}{\partial \mathbf{A}} = \left(\mathbf{B}+\mathbf{B}^{\dagger}\right)\ (\mathbf{A}\,\boldsymbol{x}-\boldsymbol{y})\ \boldsymbol{x}^{\dagger}$

Continuous Fourier transform

$\widehat{f}(\nu) = \mathcal{F}(f)(\nu) = \int_{-\infty}^{+\infty} f(t)\mathrm{e}^{-2\,\mathrm{i}\,\pi\,\nu\,t}\,\mathrm{d}t$ — forward

$f(t) = \mathcal{F}^{-1}\left(\widehat{f}\right)(t) = \int_{-\infty}^{+\infty} f(\nu)\mathrm{e}^{2\,\mathrm{i}\,\pi\,\nu\,t}\,\mathrm{d}\nu$ — inverse

$\mathcal{F}(\alpha f + \beta g) = \alpha\widehat{f} + \beta\widehat{g}$ — linearity

$\mathcal{F}(f(t-t_0)) = \mathrm{e}^{-2\,\mathrm{i}\,\pi\,\nu\,t_0}\widehat{f}(\nu)$ — shift

$\mathcal{F}(\mathrm{e}^{2\,\mathrm{i}\,\pi\,\nu_0\,t} f(t)) = \widehat{f}(\nu-\nu_0)$ — modulation

$\mathcal{F}(f(a\,t)) = \frac{1}{|a|}\widehat{f}\left(\frac{\nu}{a}\right)$ — scaling

$\mathcal{F}(f^{*})(\nu) = \widehat{f}^{*}(-\nu)$ — conjugation

$\int_{-\infty}^{+\infty}\left|\widehat{f}(\nu)\right|^2\,\mathrm{d}\nu = \int_{-\infty}^{+\infty}|f(t)|^2\,\mathrm{d}t$ — Plancherel-Parseval

$\widehat{f}(0) = \int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t$ — integration

$\mathcal{F}\left(f^{(n)}\right) = (2\,\mathrm{i}\,\pi\,\nu)^n\ \widehat{f}(\nu)$ — differentiation

$\mathcal{F}(f * g) = \widehat{f}\,\widehat{g}$ — convolution

$\mathcal{F}(f\star f) = \left|\widehat{f}\right|^2$ — autocorrelation

Discrete Fourier Transform

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)(N-1)} \end{bmatrix}$$

with $\omega = \mathrm{e}^{-\frac{-2\mathrm{i}\pi}{N}}$ the N^{th} root of unity.

$\mathbf{F}^{-1} = \frac{1}{N}\mathbf{F}^{\dagger}$ — orthogonality

$\mathbf{U} = \frac{1}{\sqrt{N}}\mathbf{F}$ — $\mathbf{U}$ is unitary

$\|\boldsymbol{x}\|_2^2 = \frac{1}{N}\|\mathbf{F}\,\boldsymbol{x}\|_2^2$ — Plancherel-Parseval

$\|\boldsymbol{x}\|_1 \leq \|\mathbf{F}\,\boldsymbol{x}\|_1 \leq N\ \|\boldsymbol{x}\|_1$

Circular convolution matrix H

$H_{0,i} = h_i$ — impulse response (PSF)

$\mathbf{H} = \mathbf{F}^{-1}\,\mathrm{diag}(\widehat{\boldsymbol{h}})\,\mathbf{F}$ — diagonalization by Fourier

$\widehat{\boldsymbol{h}} = \mathbf{F}\,\boldsymbol{h}$ — eigenvalues spectrum

Continuous probability distribution

$\boldsymbol{x} \in \mathbb{C}^N$ is a continuous random vector, it has a probability density function (pdf) $f_X(\boldsymbol{x})$ such that, for all $\mathbb{A} \subseteq \mathbb{C}^N$:

$$\Pr(\boldsymbol{x} \in \mathbb{A}) = \int_{\mathbb{A}} f_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

$\bar{\boldsymbol{x}} = \mathrm{E}(\boldsymbol{x}) = \langle\boldsymbol{x}\rangle = \int_{-\infty}^{+\infty}\boldsymbol{x}\,f_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ — Expectation (mean)

$\mathrm{E}(\alpha\,\boldsymbol{x} + \beta\,\boldsymbol{y} + \gamma) = \alpha\,\mathrm{E}(\boldsymbol{x}) + \beta\,\mathrm{E}(\boldsymbol{y}) + \gamma$ — linearity

$\mathbf{C}_{\boldsymbol{x},\boldsymbol{y}} = \mathrm{Cov}\,(\boldsymbol{x},\boldsymbol{y}) = \mathrm{E}(\boldsymbol{x}\boldsymbol{y}^{\dagger}) - \mathrm{E}(\boldsymbol{x})\,\mathrm{E}(\boldsymbol{y})^{\dagger}$ — Cross-covariance

$\mathbf{C}_{\boldsymbol{x}} = \mathrm{Cov}\,(\boldsymbol{x},\boldsymbol{x}) = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}^{\dagger}) - \mathrm{E}(\boldsymbol{x})\,\mathrm{E}(\boldsymbol{x})^{\dagger}$ — Covariance

$\mathbf{C}_{\boldsymbol{x}}$ is an Hermitian matrix of size $N \times N$.

Independence & Uncorrelation

$\Pr(X \mid Y) = \Pr(X)$ — Independence

$f_{X,Y}(\boldsymbol{x},\boldsymbol{y}) = f_X(\boldsymbol{x})\,f_Y(\boldsymbol{y})$ — Independence

$\mathrm{Cov}\,(\alpha\,\boldsymbol{x} + \beta\,\boldsymbol{y} + \gamma) = \alpha^2\,\mathrm{Cov}\,(\boldsymbol{x}) + \beta^2\,\mathrm{Cov}\,(\boldsymbol{y})$ — Uncorrelation

$\mathrm{Cov}\,(\boldsymbol{x},\boldsymbol{y}) = 0$ — Uncorrelation

independence $\Rightarrow$ uncorrelation — uncorrelation $\not\Rightarrow$ independence

$\Pr(X \mid Y) = \frac{\Pr(Y \mid X)\,\Pr(X)}{\Pr(Y)}$ — Bayes rule

Convexity

$f : \mathbb{C}^N \to \mathbb{R}$ is strictly convex if and only if:

$$f(\lambda\,\boldsymbol{x} + (1-\lambda)\,\boldsymbol{y}) < \lambda\,f(\boldsymbol{x}) + (1-\lambda)\,f(\boldsymbol{y}) \quad \forall\,\boldsymbol{x},\boldsymbol{y},\ \lambda \in]0,1[\,.$$

$\boldsymbol{x}^{+} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$ — global minimum

$\boldsymbol{p} \in \partial f(\boldsymbol{x}) \Leftrightarrow f(\boldsymbol{x}) - f(\boldsymbol{x}') \geq \langle\boldsymbol{p}, \boldsymbol{x}-\boldsymbol{x}'\rangle\ ,\forall\boldsymbol{x}'$ — subgradient

Gradient descent

$f : \mathbb{C}^N \to \mathbb{R}$ is convex and differentiable with Lipschitz gradient L:

$$\|\nabla\,\boldsymbol{x} - \nabla\,\boldsymbol{y}\|_2 \leq L\ \|\boldsymbol{x}-\boldsymbol{y}\|_2$$

The following sequence converges toward a minimizer of f in $O(1/k)$:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \gamma\nabla f(\boldsymbol{x}^{(k)}) \text{ with } \gamma \in]0, 1/L]$$

Newton's method

$f : \mathbb{C}^N \to \mathbb{R}$ is convex and twice differentiable, the sequence:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \left(\nabla^2 f(\boldsymbol{x}^{(k)})\right)^{-1}\nabla f(\boldsymbol{x}^{(k)})$$

converges toward a minimizer of f in $O(1/k^2)$:

Projection

$\mathbf{P} \in \mathbb{C}^{N\times N}$ is a projection on a subset $\mathbb{S} \subseteq \mathbb{C}^N$ and its indicator $\imath_{\mathbb{S}}$:

$$\mathbf{P}\,\boldsymbol{x} = \arg\min_{\boldsymbol{y}}\left(\imath_{\mathbb{S}}(\boldsymbol{y}) + \tfrac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2\right) \quad \text{with } \imath_{\mathbb{S}}(\boldsymbol{x}) \begin{cases} 0 & \text{if } \boldsymbol{x} \in \mathbb{S}\,, \\ +\infty & \text{otherwise}\,. \end{cases}$$

$\mathbf{P}\,\mathbf{P} = \mathbf{P}\,.$ — idempotent

$\|\mathbf{P}\,\boldsymbol{x} - \mathbf{P}\,\boldsymbol{y}\|_2 \leq \|\boldsymbol{x}-\boldsymbol{y}\|_2$ — $\mathbb{S}$ is convex $\Rightarrow \mathbf{P}$ is non-expansive

$\boldsymbol{x}^{(k+1)} = \mathbf{P}\left(\boldsymbol{x}^{(k)} - \gamma\nabla f(\boldsymbol{x}^{(k)})\right)$ — projected gradient descent

Proximity operator

f: a lower semi-continuous convex function, its proximity operator is:

$$\mathrm{prox}_f(\boldsymbol{y}) = \arg\min_{\boldsymbol{y}}\left(f(\boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2\right)$$

$\boldsymbol{p} = \mathrm{prox}_f(\boldsymbol{x}) \Leftrightarrow \boldsymbol{x} - \boldsymbol{p} \in \partial f(\boldsymbol{p})$ with $\partial f(\boldsymbol{p})$ the subgradient of f

$\boldsymbol{x}^{+} = \mathrm{prox}_f(\boldsymbol{x}^{+}) \Leftrightarrow \boldsymbol{x}^{+} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$