11_Projected_Gradient_Descent

Martin Reißel

27. Juni 2022

Inhaltsverzeichnis

1 Projected Gradient-Descent **1**
1.1 Überblick . . . 1
1.2 Grundlagen . . . 1
1.3 Konvergenz . . . 6
1.3.1 Vorüberlegungen . . . 6
1.3.2 Lipschitz-Stetigkeit . . . 6
1.3.3 L-Glattheit . . . 7
1.3.4 μ-Konvexität . . . 9
1.4 Projektionsoperatoren . . . 10
1.4.1 Kugel in der 2-Norm . . . 10
1.4.2 Nicht negativer Kegel . . . 10
1.4.3 Kugel in der 1-Norm . . . 11
1.5 Beispiel Tomographie . . . 18
1.6 Zusammenfassung . . . 20

1 Projected Gradient-Descent

1.1 Überblick

Bisher haben wir Gradient-Descent für nicht restringierte Optimierungsprobleme untersucht. In diesem Abschnitt modifizieren wir Gradient-Descent so, dass wir es auch auf eine bestimmte Klasse von restringierten Optimierungsproblemen anwenden können. Dazu kombinieren wir Gradient-Descent mit einer geeigneten Projektion auf die zulässige Parametermenge $X \subset \mathbb{R}^d$.

1.2 Grundlagen

In den Beispielen oben haben wir bereits Projected-Gradient-Descent benutzt, um für $f \in C^1(\mathbb{R}^d)$ restringierte Optimierungsproblem näherungsweise zu lösen:

$$
\begin{aligned}
x_* &= \operatorname{argmin}_{x \in X} f(x), \quad X \subset \mathbb{R}^d \\
y_{t+1} &= x_t - \gamma_t f'(x_t) \\
x_{t+1} &= \operatorname{argmin}_{x \in X} \|x - y_{t+1}\|
\end{aligned}
$$

Ist x_{t+1} überhaupt wohldefiniert (Existenz, Eindeutigkeit)?

Satz: Ist V eine Hilbert-Raum, $\emptyset \neq X \subset V$ konvex und abgeschlossen, dann existiert genau ein $y \in X$ (die Projektion von x auf X) mit

$$\|x - y\| = \inf_{z \in X} \|z - y\|.$$

Für $x \in V$ ist $y \in X$ die zugehörige Projektion, genau dann wenn

$$(x - y, z - y) \leq 0 \quad \forall z \in X.$$

Beweis:

- Existenz von y:
  - sei $y_i \in X$ eine Folge mit
    $$\lim_{i \to \infty} \|x - y_i\| = \inf_{z \in X} \|x - z\| =: d$$
  - wegen $\|v\|^2 = (v, v)$ gilt die Parallelogrammgleichung
    $$\|u + v\|^2 + \|u - v\|^2 = 2\big(\|u\|^2 + \|v\|^2\big)$$
    also
    $$\|u - v\|^2 = 2\big(\|u\|^2 + \|v\|^2\big) - \|u + v\|^2$$
  - setzen wir nun
    $$u = y_i - x, \quad v = y_j - x$$
    dann folgt
    $$\|y_i - y_j\|^2 = 2\big(\|y_i - x\|^2 + \|y_j - x\|^2\big) - \|y_i + y_j - 2x\|^2$$
  - wegen
    $$\|y_i + y_j - 2x\|^2 = 4\Big\|\underbrace{\frac{y_i + y_j}{2}}_{\in X} - x\Big\|^2 \geq 4d^2$$
    gilt
    $$\|y_i - y_j\|^2 \leq 2\big(\|y_i - x\|^2 + \|y_j - x\|^2\big) - 4d^2$$
  - andererseits gibt es für alle $\varepsilon > 0$ ein $N_\varepsilon \in \mathbb{N}$ mit
    $$\|y_k - x\|^2 \leq d^2 + \varepsilon \quad \forall k \geq N_\varepsilon$$
    und somit
    $$\|y_i - y_j\|^2 \leq 2\big(d^2 + \varepsilon + d^2 + \varepsilon\big) - 4d^2 = 4\varepsilon \quad \forall i, j \geq N_\varepsilon$$
  - damit ist y_i eine Cauchy-Folge
  - da V vollständig ist gibt es ein $y \in V$ mit $y_i \xrightarrow{i \to \infty} y$
  - wegen $y_i \in X$, X abgeschlossen, ist dann auch $y \in X$
- Eindeutigkeit von y:
  - für $y, \tilde{y} \in X$ gelte
    $$d = \inf_{z \in X} \|z - x\| = \|y - x\| = \|\tilde{y} - x\|$$
  - da X konvex ist gilt
    $$\bar{y} = \frac{y + \tilde{y}}{2} \in X$$

und mit der Parallelogrammgleichung folgt

$$\begin{aligned}\|\bar{y}-x\|^2 &= \left\|\frac{y+\tilde{y}}{2}-x\right\|^2\\ &= \frac{1}{4}\|\underbrace{y-x}_{u}+\underbrace{\tilde{y}-x}_{v}\|^2\\ &= \frac{1}{4}\|u+v\|^2\\ &= \frac{1}{2}(\|u\|^2+\|v\|^2)-\frac{1}{4}\|u-v\|^2\\ &= \frac{1}{2}(\|y-x\|^2+\|\tilde{y}-x\|^2)-\frac{1}{4}\|y-\tilde{y}\|^2\\ &= d^2-\frac{1}{4}\|y-\tilde{y}\|^2\end{aligned}$$

- für $y \neq \tilde{y}$ ist
$$\|\bar{y}-x\|^2 < d^2$$
was ein Widerspruch zu $d = \inf_{z\in X}\|z-x\|$ ist

- Ungleichung:
  - es sei $y \in X$ der eindeutige Minimierer und $d = \|y-x\| = \inf_{z\in X}\|z-x\|$:
    - da X konvex ist gilt für alle $z \in X$ und alle $\lambda \in (0,1)$
    $$(1-\lambda)y+\lambda z = y+\lambda(z-y) \in X$$
    - damit folgt
    $$\begin{aligned}\|y-x\|^2 = d^2 &\leq \|y+\lambda(z-y)-x\|^2\\ &= \|y-x\|^2+\lambda^2\|z-y\|^2+2\lambda(y-x,z-y)\end{aligned}$$
    also
    $$2\lambda(x-y,z-y) \leq \lambda^2\|z-y\|^2$$
    und wegen $\lambda > 0$
    $$(x-y,z-y) \leq \frac{\lambda}{2}\|z-y\|^2$$
    - durch Grenzübergang $\lambda \downarrow 0$ erhalten wir schließlich
    $$(x-y,z-y) \leq 0 \quad \forall z \in X$$
  - es sei nun $y \in X$ mit $(x-y,z-y) \leq 0\ \forall z \in X$:
    - für $z \in X$, $z \neq y$, folgt dann
    $$\begin{aligned}\|x-z\|^2 &= \|x-y+y-z\|^2\\ &= \|x-y\|^2+\underbrace{\|y-z\|^2}_{>0}\underbrace{-2(x-y,z-y)}_{\geq 0}\\ &> \|x-y\|^2\end{aligned}$$
    - somit ist y Minimierer von $\|x-z\|$, $z \in X$

□

Mit diesem Ergebnis machen die folgenden Definitionen Sinn.

Definition: Ist V eine Hilbert-Raum, $\varnothing \neq X \subset V$ konvex und abgeschlossen, dann ist

$$d_X : V \to \mathbb{R}, \quad d_X(x) = \inf_{z \in X} \|z - x\|$$

der *Abstand* von x zu X und

$$\Pi_X : V \to X, \quad \Pi_X(x) = \text{argmin}_{z \in X} \|z - x\|$$

die *Projektion* von x auf X.

Bemerkung:

- wir lassen bei d, Π den Index X weg
- Π ist wohldefiniert, wenn V vollständig und die Norm auf V strikt konvex ist
- wird die Norm von einem Skalarprodukt induziert, dann ist sie immer strikt konvex
- ist dies nicht der Fall, kann man einfache Gegenbeispiele konstruieren

Beispiel:

- betrachte
$$V = \mathbb{R}^2, \quad \|x\|_\infty = \max(|x_1|, |x_2|),$$
- $(V, \|\cdot\|_\infty)$ ist ein Banachraum, also ein vollständiger normierter Vektorraum
- für die konvexe Menge
$$X = [-1, 1] \times [-1, 1]$$
und den Punkt $x = (2, 0)^T$ erhalten wir
$$d(x) = \inf_{z \in X} \|z - x\|_\infty = \inf_{z \in X} \max(|z_1 - 2|, |z_2|)$$
- wegen $z_1 \in [-1, 1]$ ist für alle $z \in X$
$$|z_1 - 2| \geq 1,$$
s.d. $d(x) \geq 1$
- andererseits gilt für alle z mit $z_1 = 1$ und $z_2 \in [-1, 1]$
$$\|z - x\|_\infty = \max(|z_1 - 2|, |z_2|) = 1,$$
der minimale Abstand $d(x) = 1$ wird also für alle diese z angenommen
- $d(x)$ ist also nach wie vor wohldefiniert, $\Pi(x)$ wegen der fehlenden Eindeutigkeit dagegen nicht

Jetzt betrachten wir einige wichtige Eigenschaften von d und Π.

Lemma: Sind V, X, d, Π wie oben definiert, dann gilt:

- $d : V \to \mathbb{R}$ ist konvex
- $\big(y - \Pi(y), x - \Pi(y)\big) \leq 0 \quad \forall x \in X,\ \forall y \in V$
- $\|\Pi(y) - \Pi(x)\| \leq \|y - x\| \quad \forall x, y \in V$
- $\|x - \Pi(y)\|^2 + \|y - \Pi(y)\|^2 \leq \|y - x\|^2 \quad \forall x \in X,\ \forall y \in V$

Beweis:

- Teil 1:
  - $x, y \in V$, $\bar{x} = \Pi(x)$, $\bar{y} = \Pi(y)$

- $\bar{x}, \bar{y} \in X$ und da X konvex ist gilt

$$\bar{z} = (1-\lambda)\bar{x} + \lambda\bar{y} \in X \quad \forall \lambda \in [0,1]$$

- mit $u = (1-\lambda)x + \lambda y$ folgt

$$\begin{aligned} d\big((1-\lambda)x + \lambda y\big) &= \inf_{z \in X} \|z - u\| \\ &\le \|u - \bar{z}\| \\ &= \|(1-\lambda)x + \lambda y - (1-\lambda)\bar{x} - \lambda\bar{y}\| \\ &= \|(1-\lambda)(x-\bar{x}) + \lambda(y-\bar{y})\| \\ &\le (1-\lambda)\|x-\bar{x}\| + \lambda\|y-\bar{y}\| \\ &= (1-\lambda)d(x) + \lambda d(y) \end{aligned}$$

- Teil 2:
  - nach dem Satz von oben gilt für $y \in V$

$$\bar{y} \in X, \quad \|y - \bar{y}\| = \inf_{x \in X} \|y - x\| \quad \Leftrightarrow \quad (y - \bar{y}, x - \bar{y}) \le 0 \quad \forall x \in X$$

  - mit $\bar{y} = \Pi(y)$ folgt die Behauptung
- Teil 3:
  - es sei wieder $x, y \in V$, $\bar{x} = \Pi(x)$, $\bar{y} = \Pi(y)$
  - nach Teil 2 gilt

$$(x - \bar{x}, z - \bar{x}) \le 0, \quad (y - \bar{y}, z - \bar{y}) \le 0 \quad \forall z \in X$$

  - mit $z = \bar{y}$ folgt aus der ersten Ungleichung

$$(x - \bar{x}, \bar{y} - \bar{x}) \le 0 \quad \Rightarrow \quad (\bar{x} - x, \bar{x} - \bar{y}) \le 0$$

    und analog mit $z = \bar{x}$ in der zweiten Ungleichung

$$(y - \bar{y}, \bar{x} - \bar{y}) \le 0$$

  - Addition der beiden Ungleichungen liefert

$$(\bar{x} - x + y - \bar{y}, \bar{x} - \bar{y}) \le 0$$

    also

$$-(x - y, \bar{x} - \bar{y}) + \|\bar{x} - \bar{y}\|^2 \le 0$$

  - mit Cauchy-Schwarz folgt

$$\|\bar{x} - \bar{y}\|^2 \le (x - y, \bar{x} - \bar{y}) \le \|x - y\| \, \|\bar{x} - \bar{y}\|$$

    und somit

$$\|\bar{x} - \bar{y}\| \le \|x - y\|$$

- Teil 4:
  - $\forall v, w \in V$ gilt

$$\|v - w\|^2 = (v - w, v - w) = \|v\|^2 + \|w\|^2 - 2(v, w)$$

    also

$$\|v\|^2 + \|w\|^2 - \|v - w\|^2 = 2(v, w)$$

  - mit $v = x - \bar{y}$, $w = y - \bar{y}$ folgt

$$\|x - \bar{y}\|^2 + \|y - \bar{y}\|^2 - \|x - y\|^2 = 2(x - \bar{y}, y - \bar{y}) \le 0$$

□

1.3 Konvergenz

Wir werden jetzt analog zum einfachen Gradientenverfahren Konvergenzaussagen unter verschiedenen Voraussetzungen an f beweisen. Dabei werden wir sehen, dass die asymptotischen Aussagen sich nicht verändern, d.h. die Projektion auf abgeschlossene konvexe Mengen hat keinen negativen Einfluss.

1.3.1 Vorüberlegungen

Für $f \in C^1(\mathbb{R}^d)$, $X \subset \mathbb{R}^d$ konvex und abgeschlossen, wenden wir auf das restringierte Optimierungsproblem

$$x_* = \operatorname{argmin}_{x \in X} f(x)$$

Projected-Gradient-Descent

$$y_{t+1} = x_t - \gamma f'(x_t), \quad x_{t+1} = \Pi(y_{t+1})$$

mit konstanter Schrittweite γ an, wobei Π die Projektion in $\mathbb{R}^d$ auf X bezüglich $(\cdot,\cdot)_2$ ist. Wegen $y_{t+1} = x_t - \gamma f_t'$ gilt

$$\|y_{t+1} - x_*\|_2^2 = \|x_t - x_*\|_2^2 + \gamma^2 \|f_t'\|_2^2 - 2\gamma f_t'(x_t - x_*)$$

und somit

$$f_t'(x_t - x_*) = \frac{1}{2\gamma}\left(\gamma^2 \left\|f_t'\right\|_2^2 + \|x_t - x_*\|_2^2 - \|y_{t+1} - x_*\|_2^2\right).$$

Der einzige Unterschied zu Gradient-Descent ist y_{t+1} statt x_{t+1} im letzten Term auf der rechten Seite.

Um diesen Term zu bearbeiten, greifen wir auf die Eigenschaften der Projektion Π zurück. Für Π gilt

$$\|\Pi(y) - \Pi(x)\| \le \|y - x\| \quad \forall x, y \in V$$

bzw.

$$\Pi(x_*) = x_*$$

und somit

$$\|x_{t+1} - x_*\|_2 = \|\Pi(y_{t+1}) - \Pi(x_*)\|_2 \le \|y_{t+1} - x_*\|_2,$$

bzw.

$$f_t'(x_t - x_*) \le \frac{1}{2\gamma}\left(\gamma^2 \left\|f_t'\right\|_2^2 + \|x_t - x_*\|_2^2 - \|x_{t+1} - x_*\|_2^2\right).$$

Wir erhalten also fast das selbe Ergebnis wie bei Gradient-Descent, nur dass jetzt das "=" durch "≤" ersetzt wurde.

Ist f konvex, dann gilt

$$f(y) \ge f(x) + f'(x)(y - x)$$

und mit $y = x_*$, $x = x_t$

$$f_* \ge f_t + f_t'(x_* - x_t)$$

bzw.

$$\begin{aligned}
0 \le f_t - f_* \le f_t'(x_t - x_*) \\
&= \frac{1}{2\gamma}\left(\gamma^2 \left\|f_t'\right\|_2^2 + \|x_t - x_*\|_2^2 - \|y_{t+1} - x_*\|_2^2\right) \\
&\le \frac{1}{2\gamma}\left(\gamma^2 \left\|f_t'\right\|_2^2 + \|x_t - x_*\|_2^2 - \|x_{t+1} - x_*\|_2^2\right).
\end{aligned}$$

1.3.2 Lipschitz-Stetigkeit

$f \in C^1(\mathbb{R}^d)$ sei konvex und zusätzlich Lipschitz-Stetigkeit mit Konstante L_f, also

$$\|f(y) - f(x)\| \le L_f \|y - x\|$$

was äquivalent ist zu

$$\|f'(x)\| \le L_f.$$

Wie bei Gradient-Descent folgt

$$\begin{aligned}\sum_{t=0}^{T-1}(f_t - f_*) &\leq \sum_{t=0}^{T-1} f_t'(x_t - x_*)\\ &\leq \frac{\gamma}{2}TL_f^2 + \frac{1}{2\gamma}\big(\underbrace{\|x_0 - x_*\|_2^2}_{e_0^2} - \underbrace{\|x_T - x_*\|_2^2}_{\geq 0}\big)\\ &\leq \frac{\gamma T L_f^2}{2} + \frac{e_0^2}{2\gamma}.\end{aligned}$$

Satz: $f : \mathbb{R}^d \to \mathbb{R}$, konvex, C^1, L-stetig mit Konstante L_f und es existiere $x_* = \operatorname{argmin}_{x\in\mathbb{R}^d} f(x)$. Mit $\gamma = \frac{c}{T^\omega}$, $\omega \in (0,1)$, gilt

$$\begin{aligned}\min_{t=0,\dots,T-1}(f_t - f_*) &\leq \frac{1}{T}\sum_{t=0}^{T-1}(f_t - f_*)\\ &= \mathcal{O}\Big(\big(\frac{1}{T}\big)^{\min(\omega,1-\omega)}\Big) \quad \text{für} \quad T \to \infty.\end{aligned}$$

Die optimale Ordnung ist $\frac{1}{2}$ für $\omega = \frac{1}{2}$. Mit $e_0 = \|x_0 - x_*\|$, $\gamma = \frac{e_0}{L_f\sqrt{T}}$ gilt außerdem

$$\min_{t=0,\dots,T-1}(f_t - f_*) \leq \frac{1}{T}\sum_{t=0}^{T-1}(f_t - f_*) \leq \frac{L_f e_0}{\sqrt{T}}.$$

1.3.3 L-Glattheit

Wir leiten zuerst eine neue Version des Descent-Lemmas her. Dazu starten wir mit der L-Glattheit von f

$$f_{t+1} \leq f_t + f_t'(x_{t+1} - x_t) + \frac{L}{2}\|x_{t+1} - x_t\|_2^2$$

benutzen

$$y_{t+1} = x_t - \gamma f_t'$$

bzw.

$$f_t' = \frac{1}{\gamma}(x_t - y_{t+1})$$

und erhalten

$$f_{t+1} \leq f_t - \frac{1}{\gamma}\underbrace{(y_{t+1} - x_t)}_{u}{}^T\underbrace{(x_{t+1} - x_t)}_{v} + \frac{L}{2}\|x_{t+1} - x_t\|_2^2.$$

Wegen $2u^Tv = 2(u,v)_2 = \|u\|_2^2 + \|v\|_2^2 - \|u - v\|_2^2$ folgt

$$\begin{aligned}f_{t+1} &\leq f_t - \frac{1}{2\gamma}\big(\|y_{t+1} - x_t\|_2^2 + \|x_{t+1} - x_t\|_2^2 - \|y_{t+1} - x_{t+1}\|_2^2\big) + \frac{L}{2}\|x_{t+1} - x_t\|_2^2\\ &= f_t - \frac{1}{2\gamma}\big(\gamma^2\|f_t'\|_2^2 + \|x_{t+1} - x_t\|_2^2 - \|y_{t+1} - x_{t+1}\|_2^2\big) + \frac{L}{2}\|x_{t+1} - x_t\|_2^2\\ &= f_t - \frac{\gamma}{2}\|f_t'\|_2^2 + \frac{1}{2\gamma}\|y_{t+1} - x_{t+1}\|_2^2 + \frac{1}{2}\big(L - \frac{1}{\gamma}\big)\|x_{t+1} - x_t\|_2^2.\end{aligned}$$

Für $0 < \gamma \leq \frac{1}{L}$ ist $L - \frac{1}{\gamma} \leq 0$ und wir erhalten das folgende Ergebnis.

Descent-Lemma: Ist $f : \mathbb{R}^d \to \mathbb{R}$ L-glatt und $0 < \gamma \leq \frac{1}{L}$, dann gilt für Projected-Gradient-Descent

$$f_{t+1} \leq f_t - \frac{\gamma}{2}\|f_t'\|_2^2 + \frac{1}{2\gamma}\|y_{t+1} - x_{t+1}\|_2^2.$$

Bemerkung:

- bei Gradient-Descent war
$$f_{t+1} \le f_t - \beta \|f_t'\|_2^2, \quad \beta = \gamma\Big(1 - \frac{\gamma L}{2}\Big),$$
so dass die f_t bei $0 < \gamma < 2/L$ monoton fallend waren
- bei Projected-Gradient-Descent haben wir auf der rechten Seite den nicht-negativen Zusatzterm
$$\frac{1}{2\gamma}\|y_{t+1} - x_{t+1}\|_2^2,$$
so dass wir zunächst nichts über die Monotonie der f_t aussagen können
- man kann aber trotzdem zeigen, dass die f_t auch hier (wenn γ kurz genug ist) monoton fallen (Übung)

Aus den Vorüberlegungen wissen wir (f konvex)
$$\begin{aligned} 0 \le f_t - f_* &\le f_t'(x_t - x_*) \\ &= \frac{1}{2\gamma}\big(\gamma^2\|f_t'\|_2^2 + \|x_t - x_*\|_2^2 - \|y_{t+1} - x_*\|_2^2\big). \end{aligned}$$

Laut Descent-Lemma ist
$$\|f_t'\|_2^2 \le \frac{2}{\gamma}\Big(f_t - f_{t+1} + \frac{1}{2\gamma}\|y_{t+1} - x_{t+1}\|_2^2\Big)$$
und oben eingesetzt folgt
$$f_t - f_* \le f_t - f_{t+1} + \frac{1}{2\gamma}\big(\|y_{t+1} - x_{t+1}\|_2^2 + \|x_t - x_*\|_2^2 - \|y_{t+1} - x_*\|_2^2\big).$$

Wegen der Eigenschaft
$$\|x - \Pi(y)\|^2 + \|y - \Pi(y)\|^2 \le \|y - x\|^2 \quad \forall x \in X,\ \forall y \in V$$
der Projektion Π und $x_* \in X$ folgt mit $y = y_{t+1}$, $x = x_*$
$$\begin{aligned} \|y_{t+1} - x_{t+1}\|_2^2 &= \|y_{t+1} - \Pi(y_{t+1})\|_2^2 \\ &\le \|y_{t+1} - x_*\|_2^2 - \|x_* - \Pi(y_{t+1})\|_2^2 \\ &= \|y_{t+1} - x_*\|_2^2 - \|x_{t+1} - x_*\|_2^2 \end{aligned}$$
so dass
$$f_t - f_* \le f_t - f_{t+1} + \frac{1}{2\gamma}\big(\|x_t - x_*\|_2^2 - \|x_{t+1} - x_*\|_2^2\big).$$

Durch Summation erhalten wir
$$\sum_{t=0}^{T-1}(f_t - f_*) \le f_0 - f_T + \frac{1}{2\gamma}\big(\|x_0 - x_*\|_2^2 - \|x_T - x_*\|_2^2\big).$$

Benutzen wir jetzt noch
$$f_0 - f_T = (f_0 - f_*) - (f_T - f_*), \quad \|x_T - x_*\|_2^2 \ge 0,$$
dann ist
$$\sum_{t=1}^{T}(f_t - f_*) \le \frac{1}{2\gamma}\|x_0 - x_*\|_2^2.$$

Nach dem Descent-Lemma und der daran anschließenden Bemerkung gilt $f_{t+1} \le f_t$ und somit
$$f_T - f_* \le \frac{1}{T}\sum_{t=1}^{T}(f_t - f_*) \le \frac{1}{2\gamma T}\|x_0 - x_*\|_2^2.$$

Satz: $f : \mathbb{R}^d \to \mathbb{R}$, konvex, L-glatt mit Konstante L und es existiere $x_* = \operatorname{argmin}_{x \in X} f(x)$. Ist $0 < \gamma \le \frac{1}{L}$ dann gilt für Projected-Gradient-Descent
$$f_T - f_* \le \frac{1}{2\gamma T}\|x_0 - x_*\|_2^2 = \mathcal{O}\Big(\frac{1}{T}\Big) \quad \text{für} \quad T \to \infty.$$

1.3.4 μ-Konvexität

Ist f μ-konvex mit $\mu \geq 0$, dann gilt

$$f_* \geq f_t + f_t'(x_* - x_t) + \frac{\mu}{2}\|x_* - x_t\|_2^2$$

bzw.

$$f_t - f_* \leq f_t'(x_t - x_*) - \frac{\mu}{2}\|x_t - x_*\|_2^2.$$

Aus den Vorüberlegungen wissen wir

$$f_t'(x_t - x_*) = \frac{1}{2\gamma}\left(\gamma^2\|f_t'\|_2^2 + \|x_t - x_*\|_2^2 - \|y_{t+1} - x_*\|_2^2\right).$$

Wegen

$$\|x - \Pi(y)\|^2 + \|y - \Pi(y)\|^2 \leq \|x - y\|^2 \quad \forall x \in X \quad \forall y \in V$$

folgt mit $y = y_{t+1}$, $x = x_*$, $\Pi(y_{t+1}) = x_{t+1}$

$$\|x_* - x_{t+1}\|_2^2 + \|y_{t+1} - x_{t+1}\|_2^2 \leq \|x_* - y_{t+1}\|_2^2,$$

also

$$-\|y_{t+1} - x_*\|_2^2 \leq -\|x_{t+1} - x_*\|_2^2 - \|y_{t+1} - x_{t+1}\|_2^2.$$

Oben eingesetzt erhalten wir nun

$$\begin{aligned}f_t - f_* &\leq f_t'(x_t - x_*) - \frac{\mu}{2}\|x_t - x_*\|_2^2\\ &\leq \frac{1}{2\gamma}\left(\gamma^2\|f_t'\|_2^2 + \|x_t - x_*\|_2^2 - \|x_{t+1} - x_*\|_2^2 - \|y_{t+1} - x_{t+1}\|_2^2\right)\\ &\quad - \frac{\mu}{2}\|x_t - x_*\|_2^2,\end{aligned}$$

also

$$\begin{aligned}2\gamma(f_t - f_*) \leq{}& \gamma^2\|f_t'\|_2^2 + (1 - \gamma\mu)\|x_t - x_*\|_2^2\\ &- \|x_{t+1} - x_*\|_2^2 - \|y_{t+1} - x_{t+1}\|_2^2\end{aligned}$$

bzw.

$$\begin{aligned}\|x_{t+1} - x_*\|_2^2 \leq{}& 2\gamma(f_* - f_t) + \gamma^2\|f_t'\|_2^2 - \|y_{t+1} - x_{t+1}\|_2^2\\ &+ (1 - \gamma\mu)\|x_t - x_*\|_2^2.\end{aligned}$$

Die ersten drei Terme auf der rechten Seite schätzen wir wieder mit dem Descent-Lemma ab. Wegen

$$f_{t+1} \leq f_t - \frac{\gamma}{2}\|f_t'\|_2^2 + \frac{1}{2\gamma}\|y_{t+1} - x_{t+1}\|_2^2$$

ist

$$f_* - f_t \leq f_{t+1} - f_t \leq -\frac{\gamma}{2}\|f_t'\|_2^2 + \frac{1}{2\gamma}\|y_{t+1} - x_{t+1}\|_2^2$$

und somit

$$\|x_{t+1} - x_*\|_2^2 \leq (1 - \gamma\mu)\|x_t - x_*\|_2^2.$$

Analog zu Gradient-Descent haben wir damit folgendes Ergebnis gezeigt.

Satz: $f : \mathbb{R}^d \to \mathbb{R}$, μ-konvex mit $\mu > 0$, L-glatt mit Konstante L und es existiere $x_* = \operatorname{argmin}_{x \in X} f(x)$. Für $0 < \gamma \leq \frac{1}{L}$ folgt

$$\|x_{t+1} - x_*\|_2^2 \leq \rho\|x_t - x_*\|_2^2$$

und

$$f_T - f_* \leq \frac{L}{2}\rho^T\|x_0 - x_*\|_2^2$$

mit

$$\rho = 1 - \gamma\mu \in [0, 1).$$

1.4 Projektionsoperatoren

In den vorherigen Abschnitten haben wir gesehen, dass wir für Projected-Gradient-Descent exakt die selben asymptotischen Resultate erhalten wie bei gewöhnlichem Gradient-Descent. Das sieht zunächst sehr angenehm aus, für den praktischen Einsatz muss man allerdings bedenken, dass in jedem Schritt eine Projektion

$$x_{t+1} = \Pi_X(y_{t+1}), \quad \Pi_X(x) = \operatorname{argmin}_{z \in X} \|z - x\|$$

auf die abgeschlossene, konvexe Menge X zu berechnen ist.

Das Berechnen von $\Pi(x)$ ist also selbst ein restringiertes konvexes Optimierungsproblem mit einer "einfachen" konvexen Zielfunktion $x \to \|z - x\|$. Für allgemeine abgeschlossene und konvexe Menge X kann die Lösung dieses Problems extrem schwierig werden, für spezielle X, die in der Praxis häufig auftauchen, kann man Π_X dagegen explizit angeben.

Wir untersuchen einige $X \subset \mathbb{R}^d$ nun genauer. Dabei betrachten wir jeweils Projektionen bezüglich des euklidischen Skalarprodukts $(\cdot, \cdot)_2$.

1.4.1 Kugel in der 2-Norm

- wir betrachten
$$X = \bar{B}_r(0)_{\|\cdot\|_2} = \left\{x \mid x \in \mathbb{R}^d,\ \|x\|_2 \le r\right\}$$
- X ist offensichtlich konvex und abgeschlossen
- die Projektion auf X ist
$$\Pi_X(x) = \begin{cases} x & \|x\|_2 \le r \\ \frac{x}{\|x\|_2} r & \|x\|_2 > r \end{cases}$$
- mit $\bar{x} = \Pi_X(x)$ ist zu zeigen, dass
$$(x - \bar{x}, z - \bar{x})_2 \le 0 \quad \forall z \in X$$
- für $x \in X$ ist $\bar{x} = x$, also
$$(x - \bar{x}, z - \bar{x})_2 = (x - x, z - x)_2 = 0 \quad \forall z \in X$$
- für $x \notin X$, also $\|x\|_2 > r$, ist $\bar{x} = \frac{x}{\|x\|_2} r$ und somit
$$\begin{aligned} (x - \bar{x}, z - \bar{x})_2 &= \left(x - \frac{x}{\|x\|_2} r, z - \frac{x}{\|x\|_2} r\right)_2 \\ &= (x, z)_2 + r^2 - \left(x, \frac{x}{\|x\|_2} r\right)_2 - \left(\frac{x}{\|x\|_2} r, z\right)_2 \\ &= (x, z)_2 + r^2 - \|x\|_2 r - \frac{r}{\|x\|_2}(x, z)_2 \\ &= \underbrace{\left(1 - \frac{r}{\|x\|_2}\right)}_{>0}(x, z)_2 + r^2 - r\|x\|_2 \\ &\le \left(1 - \frac{r}{\|x\|_2}\right)\|x\|_2 \underbrace{\|z\|_2}_{\le r} + r^2 - r\|x\|_2 \\ &\le r\|x\|_2 - r^2 + r^2 - r\|x\|_2 \\ &= 0 \end{aligned}$$

1.4.2 Nicht negativer Kegel

- der Kegel
$$X = \left\{x \mid x \in \mathbb{R}^d,\ x_i \ge 0,\ i = 1, \dots, d\right\}$$
ist konvex und abgeschlossen

- die Projektion auf X ist definiert durch

$$\bar{x} = \Pi_X(x), \quad \bar{x}_i = \begin{cases} x_i & x_i \geq 0 \\ 0 & x_i < 0 \end{cases}$$

- für $z \in X$ folgt dann

$$\begin{aligned}(x - \bar{x}, z - \bar{x})_2 &= \sum_{i=1}^{d} (x_i - \bar{x}_i)(z_i - \bar{x}_i) \\ &= \sum_{x_i<0} \underbrace{x_i}_{<0} \underbrace{z_i}_{\geq 0} \\ &\leq 0\end{aligned}$$

1.4.3 Kugel in der 1-Norm

- wir betrachten

$$X_r = B_r(0)_{\|\cdot\|_1} = \left\{x \;\middle|\; x \in \mathbb{R}^d,\ \|x\|_1 = \sum_{i=1}^{d} |x_i| \leq r\right\}$$

- X_r ist damit konvex und abgeschlossen und die Projektion Π auf X_r bezüglich der Norm $\|\cdot\|_2$ ist wohldefiniert:

$$\Pi(v) = \operatorname{argmin}_{\|y\|_1 \leq r} \|v - y\|_2$$

- wegen $\Pi(v) = v$ für $\|v\|_1 \leq r$ reicht es, wenn wir ab jetzt Vektoren

$$v \in \mathbb{R}^d, \quad \|v\|_1 > r$$

  betrachten
- für diese gilt $\|\Pi(v)\|_1 = r$, d.h. die Projektion liegt auf dem Rand von X_r:
  - Annahme: $\|\Pi(v)\|_1 < r$
  - wegen $\|v\|_1 > r > \|\Pi(v)\|_1$ gilt

$$\lambda = \frac{r - \|\Pi(v)\|_1}{\|v\|_1 - \|\Pi(v)\|_1} \in (0, 1)$$

  - für

$$x = (1 - \lambda)\Pi(v) + \lambda v$$

    erhalten wir

$$\begin{aligned}\|x\|_1 &\leq (1 - \lambda)\|\Pi(v)\|_1 + \lambda\|v\|_1 \\ &= \|\Pi(v)\|_1 + \lambda\big(\|v\|_1 - \|\Pi(v)\|_1\big) \\ &= \|\Pi(v)\|_1 + \frac{r - \|\Pi(v)\|_1}{\|v\|_1 - \|\Pi(v)\|_1}\big(\|v\|_1 - \|\Pi(v)\|_1\big) \\ &= r\end{aligned}$$

  - andererseits gilt

$$x - v = (1 - \lambda)\big(\Pi(v) - v\big), \quad \lambda \in (0, 1)$$

    und somit

$$\|x - v\|_2 = (1 - \lambda)\|\Pi(v) - v\|_2 < \|\Pi(v) - v\|_2,$$

    was ein Widerspruch zu

$$\Pi(v) = \operatorname{argmin}_{\|y\|_1 \leq r} \|v - y\|_2$$

    ist

- wir werden jetzt die Menge der zu betrachtenden Vektoren v weiter einschränken
- ist $L : \mathbb{R}^d \to \mathbb{R}^d$ eine lineare Abbildung mit

$$\|Lw\|_1 = \|w\|_1, \quad \|Lw\|_2 = \|w\|_2 \quad \forall w \in \mathbb{R}^d$$

dann ist L umkehrbar und es gilt für $x = \Pi(w)$

$$Lx = \Pi(Lw)$$

bzw.

$$x = L^{-1}\Pi(Lw)$$

denn

$$\begin{aligned}
\|Lw - Lx\|_2 &= \|L(w - x)\|_2 \\
&= \|w - x\|_2 \\
&= \min_{\|y\|_1 \le r} \|w - y\|_2 \\
&= \min_{\|y\|_1 \le r} \|Lw - Ly\|_2 \\
&= \min_{\|L^{-1}z\|_1 \le r} \|Lw - z\|_2 \\
&= \min_{\|z\|_1 \le r} \|Lw - z\|_2
\end{aligned}$$

- die Abbildung

$$Sw = (s_1 w_1, \ldots, s_d w_d)^T, \quad s_i = \pm 1$$

besitzt diese Eigenschaften:

  - für $v \in \mathbb{R}^d$,

$$s_i = \begin{cases} 1 & v_i \ge 0 \\ -1 & v_i < 0 \end{cases},$$

  folgt dann $Sv = (|v_1|, \ldots, |v_d|)^T = |v|$ (komponentenweise) und

$$x = \Pi(v) = S^{-1}\Pi(Sv) = S^{-1}\Pi(|v|),$$

  so dass es genügt, die Projektion für Vektoren v mit $v_i \ge 0$ zu betrachten

- die Abbildung

$$Pw = (w_{\pi(1)}, \ldots, w_{\pi(d)})^T$$

wobei π eine beliebige Permutation der Indizes $1, \ldots, d$ ist, besitzt ebenfalls die oben beschriebenen Eigenschaften:

  - für $v \in \mathbb{R}^d$ betrachten wir jetzt eine Permutation π, die die Komponenten von v absteigend sortiert, also

$$v_{\pi(1)} \ge v_{\pi(2)} \ge \ldots \ge v_{\pi(d)}$$

  und erhalten

$$x = \Pi(v) = P^{-1}\Pi(Pv),$$

  so dass es genügt, die Projektion für Vektoren v zu betrachten, deren Komponenten absteigend sortiert sind

- wir müssen jetzt also nur noch Vektoren $v \in \mathbb{R}^d$ betrachten mit

$$\|v\|_1 > r, \quad v_1 \ge v_2 \ge \ldots \ge v_d \ge 0$$

und wissen bereits, dass

$$\|\Pi(v)\|_1 = r$$

ist

- unter diesen Voraussetzungen gilt für $x = \Pi(v)$ auch $x_i \geq 0$:
    - Annahme: es existiere ein $x_k < 0$
    - mit
      $$y = (x_1, \ldots, x_{k-1}, -x_k, x_{k+1}, \ldots, x_d)^T$$
      ist $\|y\|_1 = \|x\|_1 = r$ und wegen $x_k < 0 \leq v_k$
      $$\begin{aligned} \|v - y\|_2^2 &= \sum_{i=1}^{d} (v_i - y_i)^2 \\ &= (v_k + x_k)^2 + \sum_{i \neq k} (v_i - x_i)^2 \\ &< (v_k - x_k)^2 + \sum_{i \neq k} (v_i - x_i)^2 \\ &= \|v - x\|_2^2 \end{aligned}$$
      was ein Widerspruch zu $x = \Pi(v)$ ist
- somit folgt unter diesen Voraussetzungen an v
  $$\begin{aligned} \Pi(v) &= \operatorname{argmin}_{x \in Y_r} \|v - x\|_2, \\ Y_r &= \left\{ x \;\middle|\; x \in \mathbb{R}^d,\ x_i \geq 0,\ i = 1, \ldots, d,\ \sum_{i=1}^{d} x_i = r \right\}, \end{aligned}$$
  wobei Y_r wieder konvex ist
- wir zeigen jetzt, dass ein eindeutiger Index $p \in \{1, \ldots, d\}$ existieren muss mit $x_i > 0$ für $i \leq p$ und $x_i = 0$ für $i > p$:
    - $x = \Pi(v)$ minimiert die konvexe differenzierbare Funktion $f(y) = \frac{1}{2}\|y - v\|_2^2$ über der konvexen Menge Y_r, weshalb
      $$0 \leq f'(x)(y - x) = (x - v)^T (y - x) \quad \forall y \in Y_r$$
      gelten muss, also
      $$0 \leq \sum_{i=1}^{d} (x_i - v_i)(y_i - x_i) \quad \forall y \in Y_r$$
    - nach oben wissen wir, dass $\|x\|_1 = r > 0$, $x_i \geq 0$, $i = 1, \ldots, d$, ist, d.h. es existiert ein Index k mit $x_k > 0$
    - nehmen wir nun an, dass ein Index j existiert mit $x_j = 0$ und $x_{j+1} > 0$
    - definieren wir für $\varepsilon > 0$
      $$y = (x_1, \ldots, x_{j-1}, \varepsilon, x_{j+1} - \varepsilon, x_{j+2}, \ldots, x_d)^T$$
      dann ist $y_i \geq 0$, $i = 1, \ldots, d$, $\|y\|_1 = \|x\|_1 = r$, also $y \in Y_r$, und
      $$\begin{aligned} \sum_{i=1}^{d} (x_i - v_i)(y_i - x_i) &= (x_j - v_j)(y_j - x_j) + (x_{j+1} - v_{j+1})(y_{j+1} - x_{j+1}) \\ &= -v_j \varepsilon - (x_{j+1} - v_{j+1})\varepsilon \\ &= \varepsilon(\underbrace{v_{j+1} - v_j}_{\leq 0} - \underbrace{x_{j+1}}_{>0}) \\ &< 0 \end{aligned}$$
      was ein Widerspruch zu
      $$0 \leq \sum_{i=1}^{d} (x_i - v_i)(y_i - x_i) \quad \forall y \in Y_r$$
      ist

    - damit kann der Fall $x_j = 0$, $x_{j+1} > 0$ nicht auftreten
    - wegen $\|x\|_1 = r > 0$ muss $x \neq 0$ sein, weshalb es einen Index p gibt mit $x_i > 0$ für $i \le p$ und $x_i = 0$ für $i > p$
- für $x = x(p) = \Pi(v)$ gilt

$$x_i = \begin{cases} v_i - c_p > 0 & i = 1, \dots, p \\ 0 & i = p+1, \dots, d \end{cases},$$

$$c_p = \frac{1}{p}\Big(\sum_{i=1}^{p} v_i - r\Big)$$

    - Annahme: es existiere ein Index $j < p$ mit $x_j - v_j \neq x_{j+1} - v_{j+1}$
        - ohne Einschränkung können wir $x_j - v_j < x_{j+1} - v_{j+1}$ annehmen
        - wegen $j < p$ ist $x_j > 0$, so dass für $0 < \varepsilon \le x_j$

$$y = (x_1, \dots, x_{j-1}, x_j - \varepsilon, x_{j+1} + \varepsilon, x_{j+2}, \dots, x_d)^T \in Y_r$$

        - analog zum letzten Abschnitt erhalten wir

$$\begin{aligned}(x-v)^T(y-x) &= (x_j - v_j)(y_j - x_j) + (x_{j+1} - v_{j+1})(y_{j+1} - x_{j+1}) \\ &= (x_j - v_j)\varepsilon - (x_{j+1} - v_{j+1})\varepsilon \\ &< 0\end{aligned}$$

        was erneut ein Widerspruch zu

$$0 \le (x-v)^T(y-x) \quad \forall y \in Y_r$$

        ist
    - damit ist also

$$x_i = \begin{cases} v_i - c_p > 0 & i = 1, \dots, p \\ 0 & i = p+1, \dots, d \end{cases}$$

    - wegen $x \in Y_r$ gilt

$$\begin{aligned} r = \|x\|_1 = \sum_{i=1}^{d} |x_i| &= \sum_{i=1}^{p} x_i \\ &= \sum_{i=1}^{p} (v_i - c_p) \\ &= \Big(\sum_{i=1}^{p} v_i\Big) - pc_p \end{aligned}$$

    und somit

$$c_p = \frac{1}{p}\Big(\sum_{i=1}^{p} v_i - r\Big)$$

- zum Schluss muss jetzt noch der Index p bestimmt werden
- es gilt $\Pi(v) = x(p_*)$ für

$$p_* = \text{argmax}_{p \in \{1,\dots,d\}} \{v_p - c_p > 0\}$$

    - für jedes $p \in \{1, \dots, d\}$ ist

$$x_i = \begin{cases} v_i - c_p > 0 & i = 1, \dots, p \\ 0 & i = p+1, \dots, d \end{cases},$$

$$c_p = \frac{1}{p}\Big(\sum_{i=1}^{p} v_i - r\Big)$$

eindeutige Lösung von

$$\operatorname{argmin}_{z\in Z_p}\|z-v\|_2^2$$

mit

$$Z_p = \Big\{z \;\Big|\; z\in\mathbb{R}^d, z_i \geq 0,\; z_i = 0\ \forall i > p,\; \|z\|_1 = \sum_{i=1}^{p} z_i = r\Big\} \subset Y_r,$$

denn Z_p ist abgeschlossen und konvex, $f(z) = \|z-v\|_2^2$ ist konvex und für alle $z \in Z_p$ gilt

$$\begin{aligned} f'(x)(z-x) &= (x-v)^T(z-x) \\ &= \sum_{i=1}^{p}(x_i-v_i)(z_i-x_i) \\ &= -c_p\sum_{i=1}^{p}(z_i-x_i) \\ &= 0 \end{aligned}$$

- Annahme: $p < p_*$
  - dann ist

    $$\begin{aligned} Z_p \ni x(p) = \Pi(v) &= \operatorname{argmin}_{y\in Y_r}\|y-v\|_2^2 \\ &= \operatorname{argmin}_{z\in Z_p}\|z-v\|_2^2 \end{aligned}$$

    bzw.

    $$Z_{p_*} \ni x(p_*) = \operatorname{argmin}_{z\in Z_{p_*}}\|z-v\|_2^2$$

  - nach Definition ist $Z_p \subset Z_{p_*} \subset Y_r$ und alle Mengen sind konvex und abgeschlossen, so dass alle Minima eindeutig sind
  - somit ist $x(p) = x(p_*)$
  - nach Konstruktion von $x(p), x(p_*)$ gilt wegen $p < p_*$

    $$x(p)_{p+1} = 0, \quad x(p_*)_{p+1} \neq 0,$$

    was zu einem Widerspruch führt

- zusammengefasst erhalten wir damit folgenden Algorithmus zur Berechnung von $x = \Pi(v)$:
  - $v \in \mathbb{R}^d$ gegeben
  - ist $\|v\|_1 \leq r$ dann ist $x = v$
  - $\|v\|_1 > r$:
    - setze

      $$w = PSv,$$

      d.h. w enthält die Beträge der Komponenten von v absteigend sortiert
    - bestimme

      $$p_* = \operatorname{argmax}_{p\in\{1,\ldots,d\}}\Big\{w_p - \frac{1}{p}\Big(\sum_{i=1}^{p} w_i - r\Big) > 0\Big\}$$

    - berechne y als

      $$y_i = \begin{cases} w_i - c_{p_*} > 0 & i = 1,\ldots,p_* \\ 0 & i = p_*+1,\ldots,d \end{cases},$$

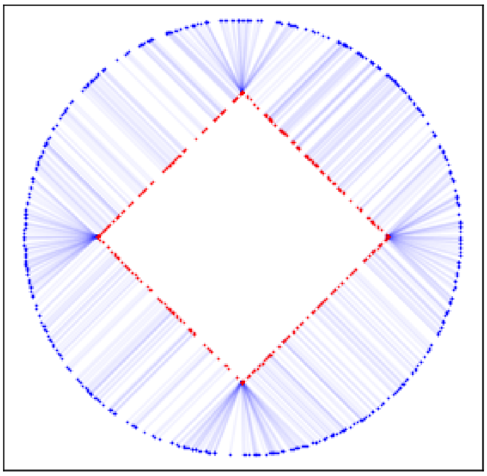
      $$c_p = \frac{1}{p_*}\Big(\sum_{i=1}^{p_*} w_i - r\Big)$$

- damit folgt

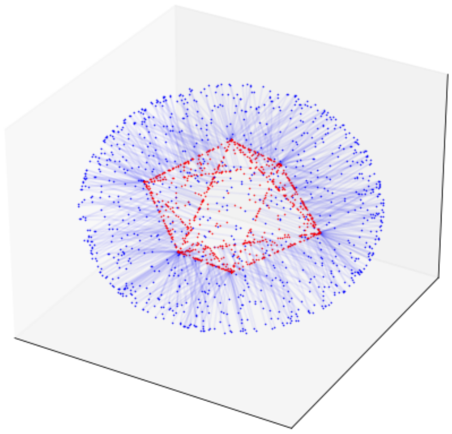
$$x = S^{-1}P^{-1}y$$

- eine Anwendung in $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ liefert folgende Resultate

```
%matplotlib inline

import numpy as np
import matplotlib.pyplot as plt

def l1pro(v, r = 1.0):
    # Projektion auf L1-Kugel mit Radius r bzgl. eukl. Skalarpr.

    if np.linalg.norm(v, ord=1) > r:
        d = v.shape[0]

        # S, Vorzeichen merken fuer S_inv
        s = np.ones(v.shape)
        s[v < 0] = -1

        # PS anwenden
        w  = s * v
        ii = w.argsort()[::-1]
        w  = w[ii]

        # Projektion
        p  = np.arange(d) + 1
        cp = (w.cumsum() - r)/p
        u  = w - cp
        jj = np.where(u > 0)[0].max()

        ps  = p[jj]
        cps = cp[jj]

        y = np.zeros(v.shape)
        y[:ps] = w[:ps] - cps

        # P_inv und S_inv anwenden
        x = np.zeros(y.shape)
        # P_inv
        x[ii] = y
        # S_inv = S
        x = s*x
    else:
        x = v

    return x


np.random.seed(17)
plt.figure(figsize=(5,5))

d = 2
n = 500
l = 1.0
r = 2.0

for k in range(n):
    w = np.random.randn(d)
    w = w/np.linalg.norm(w)*(l+r)

    x = l1pro(w, r)

    plt.plot(*w, 'b.', markersize = 1)
    plt.plot(*np.c_[x, w], 'b', alpha = 0.05)
    plt.plot(*x, 'r.', markersize = 1)

plt.axis('equal')
plt.gca().set_xticks([])
plt.gca().set_yticks([]);
```



```
#%matplotlib notebook
%matplotlib inline

plt.figure(figsize=(8,8))
plt.axes(projection='3d')
ax = plt.gca()

np.random.seed(17)

d = 3
n = 1000
l = 1.0
r = 2.0

for k in range(n):
    w = np.random.randn(d)
    w = w/np.linalg.norm(w)*(l + r)

    x = l1pro(w, r)

    ax.plot(*w, 'b.', markersize = 1)
    ax.plot(*np.c_[x, w], 'b', alpha = 0.05)
    ax.plot(*x, 'r.', markersize = 1)

#plt.axis('equal')
lr = l + r

ax.set_xlim(-lr, lr)
ax.set_ylim(-lr, lr)
ax.set_zlim(-lr, lr)
ax.set_xticks([])
ax.set_yticks([])
ax.set_zticks([]);
```



1.5 Beispiel Tomographie

Bei den Regularisierungsverfahren hatten wir für unser Tomographieproblem unter anderem die Lasso-Methode kennen gelernt. Dort bestimmt man

$$\text{argmin}_{w\in\mathbb{R}^n} \frac{1}{2m} \|Xw - y\|_2^2 + \alpha \|w\|_1,$$

wobei $\alpha > 0$ ein vom Benutzer zu wählender Regularisierungsparameter ist.

Dieses nicht restringierte Problem kann in das restringierte Problem

$$\text{argmin}_{\|w\|_1 \leq r} \frac{1}{2m} \|Xw - y\|_2^2$$

überführt werden, wobei die Rolle des Regularisierungsparameters α jetzt von r übernommen wird:

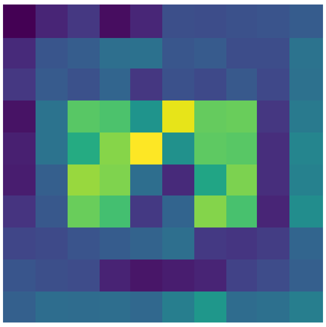
- für $r \to \infty$ ist $\|w\|_1 \leq r$ immer weniger einschränkend, man "regularisiert" immer weniger, was $\alpha \to 0$ entspricht
- für $r \to 0$ bleibt nur $w = 0$ als Lösung, d.h. man "regularisiert" maximal, was $\alpha \to \infty$ entspricht

Auf das restringierte Problem wenden wir Projected-Gradient-Descent an und benutzen dabei die Projektion auf $B_r(0)_{\|\cdot\|_1}$ aus dem letzten Abschnitt.

Zum Vergleich berechnen wir zunächst eine Rekonstruktion ohne Regularisierung.

```
%matplotlib inline

np.random.seed(17)

from DatenNotebooks.xrtomo12 import *

def plotReko(b):
    n = int(np.sqrt(b.shape[0]))

    plt.pcolor(b.reshape(n,n))
    plt.axis('equal')
    plt.axis('off')


from sklearn import linear_model
X, y, _ = tomo(delta = 0.01)

modell = linear_model.LinearRegression(fit_intercept = False)
modell.fit(X, y.ravel())
```

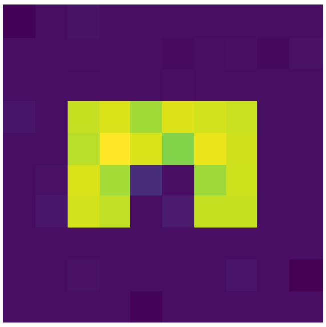
```
w = modell.coef_
plotReko(w)
```



Die Lasso-Implementierung aus Scikit-Learn liefert das folgende Ergebnis.

```
lassocv = linear_model.LassoCV(fit_intercept = False, cv = 5)

lassocv.fit(X, y)

w = lassocv.coef_
plotReko(w)

print('alpha = {:e},   ||w||_1 = {:f}'.format(lassocv.alpha_, np.linalg.norm(w, 1)))
```

```
alpha = 9.867860e-06,   ||w||_1 = 20.221146
```

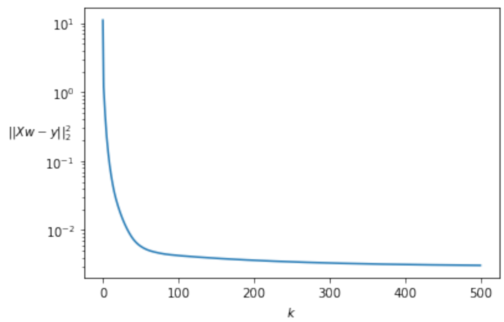


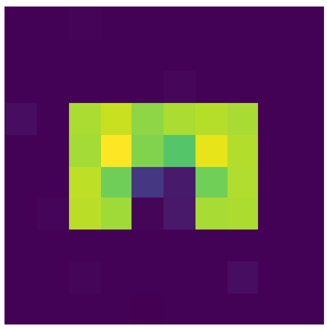
Mit Projected Gradient erhalten wir ein sehr ähnliches Resultat.

```
def l(w):
    return 0.5*((X.dot(w) - y)**2).sum()

def l1(w):
    return X.T.dot(X.dot(w) - y)

def GDp(w0, l1, pro, alpha = 1.0, nit = 10):
    w  = w0.copy()
    w = pro(w)

    ww = [w]

    for k in range(nit):
        w = w - alpha * l1(w)
        w = pro(w)
        ww.append(w)
```

```
    return ww


w0 = np.zeros(X.shape[1])

r = 20
pro1 = lambda w : l1pro(w, r)

ww = GDp(w0, l1, pro = pro1, alpha = 0.5, nit = 500)

plt.figure()
plt.semilogy(list(map(l, ww)))
plt.xlabel('$k$')
plt.ylabel('$||Xw-y||_2^2$', rotation=0)

w = ww[-1]
plt.figure()
plotReko(w)

print('r = {:f}'.format(r))
```

```
r = 20.000000
```





1.6 Zusammenfassung

Für Projected-Gradient-Descent bei restringierten Optimierungsproblemen mit $X \in \mathbb{R}^d$ abgeschlossen und konvex haben wir das selbe Konvergenzverhalten wie im nicht restringierten Fall nachgewiesen:

- f konvex und Lipschitz-stetig, $\gamma = \frac{c}{\sqrt{T}}$, $c > 0$:

$$\min_{t=0,\ldots,T-1} (f_t - f_*) \leq \varepsilon \quad \Rightarrow \quad T = \mathcal{O}\Big(\frac{1}{\varepsilon^2}\Big)$$

- f konvex und L-glatt, $0 < \gamma < \frac{2}{L}$:
$$f_T - f_* \leq \varepsilon \quad \Rightarrow \quad T = \mathcal{O}\Big(\frac{1}{\varepsilon}\Big)$$

- f μ-konvex mit $\mu > 0$ und L-glatt, $0 < \gamma \leq \frac{1}{L}$:
$$f_T - f_* \leq \varepsilon \quad \Rightarrow \quad T = \mathcal{O}\Big(\log\Big(\frac{1}{\varepsilon}\Big)\Big)$$

Das Berechnen der Projektion Π_X auf $X \in \mathbb{R}^d$ abgeschlossen und konvex kann extrem aufwendig sein. Für einige spezielle Mengen X die in der Praxis häufig auftreten lässt sich Π_X einfach berechnen.